Modèles de Fiabilité et Sciences de l'Ingénieur

Actes de la 3ème Conférence Nationale MFSI'23

Edition 6
JNF
Journée **N**ationale
de **F**iabilité

Organisée par

Unité de Recherche LaMOS

Béjaia, 19-20 Novembre 2023

LaMOS Editions, 2023

Modèles de Fiabilité
&
Sciences de l'Ingénieur

$$W(\alpha, x, \xi) \Rightarrow \max$$

Modèles de Fiabilité et Sciences de l'Ingénieur

Editeurs : Unité de Recherche LaMOS Béjaia
Adresse : Unité de Recherche LaMOS (Modélisation et Optimisation des Systèmes), Université de Béjaia, Targa Ouzamour, Béjaia, 06000 (Algérie)
Tél : 213 34 81 37 08
Tél/Fax : 213 34 81 37 09
E-Mail : unite.lamos@univ-bejaia.dz
http://www.lamos.org

ISBN : 978-9931-884-16-3

Objectifs

Cette conférence a pour objectif de rassembler les chercheurs et les professionnels de différents coins du pays pour discuter des dernières avancées et des défis actuels dans le domaine de la fiabilité et ses applications aux sciences de l'ingénieur.

Thèmes

Les 22 communications retenues par le Comité de Programme seront présentées en 04 sessions et couvrent les thématiques suivantes :

1. Fiabilité dans les sciences de l'ingénieur (mécanique, électrique, génie civil, informatique…) ;
2. Sûreté de fonctionnement des systèmes ;
3. Modèles de maintenance basés sur la fiabilité ;
4. Optimisation de la conception basée sur la fiabilité (RBDO) ;
5. Analyse de la fiabilité dans un contexte d'incertitudes ;
6. Modèles stochastiques (risques et fiabilité, systèmes à serveur non fiable, …) ;
7. Estimation non paramétrique de la fiabilité.

Comités

Président d'honneur

Pr Abedelkrim Beniaïche, Recteur de l'Université de Bejaia

Président du Colloque

Pr Djamil Aïssani, Directeur de Recherche

Comité d'Organisation

Président du Comité d'Organisation

Pr Mohamed Boualem, Directeur de l'Unité de Recherche LaMOS

Vice-Président du Comité d'Organisation

Dr Nassim Touche

Membres du Comité d'Organisation

El hassene Ait Mokhtar, Aicha Anzi, Fazia Aoudia, Mouloud Atmani, Halima Berri, Bachir Cherfaoui, Nassima Dairi, Naouel Halimi, Safia Hocine, Kamal Kabyl, Noureddine Khimoum, Samia Madi, Mohand Moktefi, Razika Sait, Sofiane Touati, Sofiane Ziani, Nabil Zougab.

Comité de Programme

Président du Comité de Programme

Pr Radouane Laggoune, Chef de l'équipe de recherche MFS

Vice-Président du Comité de Programme

Dr El Hassene Ait Mokhtar

Membres du Comité de Programme

Karim Abbas, Université de Bejaia
Karima Adel-Aissanou, Université de Bejaia
Smail, Adjabi, Université de Bejaia
Mabrouk Aib, ENP El Harrach, Alger
Amina Aissani, EURIMA, Bruxelles
Djamil Aissani, Université de Bejaia
Mourad Amad, Université de Bouira
Vladimir Anisimov,Center for Design and Analysis, Amgen - Cambridge
Younes Aoues, INSA de Rouen
Mouloud Atmani, Université de Béjaia
Kadda Baghdad Bey, EMP, Bordj-el Bahri
Aicha Bareche, Université de Bejaia
Kamel Barkaoui, CNAM Paris
Emilio Bastidas, Université de La Rochelle
Abdelhamid Becheur, Université de Bejaia
Mohand Ouamer Bibi, Université de Bejaia
Mohamed Boualem, Université de Bejaia
Louiza Bouallouche, Université de Bejaia
Ahmed Boubakeur, ENP El Harrach, Alger
Nouredine Boumahrat, Université de Boumerdes
Alexander Bochkov, JSC NIIAS, Moscou
Alaa Chateauneuf, CIDECO, Clermont-Ferrand
Ahmed Damou, A2T2, Alger
Amar Debbouche, Université de Guelma
Latefa Ghomri, Université de Tlemcen
Abdelghani Hamaz, Université de Tizi-ouzou
Bernd Heidergott, Université de Vrijl, Amsterdam
Rabia Khelif, Université d'Annaba
Abdellah Kouzou, Université de Djelfa
Karima Lagha, Université de Bejaia
Fodil Laib, Data Manger Cevital
Catalina Llado, Université des îles Baléares
Ouiza Lekadir, Université de Bejaia
Rabah Medjoudj, Université de Bejaia
Mohand Moktefi, Université de Bejaia
Assia Outamazirt, Unité de Recherche LaMOS
Mohammed Said Radjef, Université de Bejaia
Fazia Rahmoune, Université de Bejaia
Yacine Sahraoui, Université de Souk Ahras
Mustapha Réda Senouci, EMP, Bordj-el Bahri
Abdelghani Seghir, Université de Bejaia
Abdelnacer Smati, PEGAZ Engineering, Alger
János Sztrik, Université de Debrecen
Nassim Touche, Université de Bejaia
Mohamed Tounsi, Université de Bejaia
Mohand Yazid, Université de Bejaia
Iskander Zouaghi, ENP El Harrach, Alger
Nabil Zougab, Université de Bejaia

Avant – Propos

Il y a 35 ans de cela, des mathématiciens (méthodes stochastiques de la Recherche Opérationnelle, Statisticiens, spécialistes de l'Optimisation,), des spécialistes des différentes disciplines de la Science de l'Ingénieur (Electrotechniciens, mécaniciens, informaticiens, électroniciens,…) et des Gestionnaires d'Entreprise (Sonatrach, Sider, PMA/CMA, Cerhyd, Sonelgaz, EMAC, ENPEC,…) s'associaient pour « construire » ici même à Béjaia la Première Conférence Nationale ***MFSI'1988*** (*Modèles de Fiabilité et Sciences de l'Ingénieur)*. Il s'agissait de la première manifestation jamais organisée dans notre pays sur la fiabilité [voir le Compte Rendu publié par la revue ***Es-Syana – la Maintenance***, Vol. 1, INMA, Ministère de l'Industrie Lourde édition, 1988, pp. 27 – 28]**.** Dix ans plus tard, le Ministère de la Défense Nationale, à travers l'***ENITA*** (aujourd'hui l'***Ecole Militaire Polytechnique)*** avait pris le relais pour l'organisation, à Bordj-el-Bahri, de la 2ème édition [voir le Compte Rendu paru dans la revue ***MATAPLI*** de la ***SMAI*** *(Société Française des Mathématiques Appliquées et Industrielles*), n° 54, 1998, pp. 65 – 66].

Peut-on dire quelques mots sur les raisons qui font qu'en 2023 la Conférence Nationale ***M.F.S.I.*** revient dans la ville qui l'a vu naître ? Il s'agit d'abord du souci des initiateurs du projet de faire le point sur le chemin parcouru. Pas simplement en terme de développement de la discipline ou bien en terme d'applications au niveau des sciences de l'ingénieur et des différents utilisateurs (entreprises industrielles, organismes socio-économiques). On constate aujourd'hui que les efforts fournis ont abouti à la structuration d'une véritable ***Ecole Algérienne de Fiabilité***. A titre d'exemple, on observe que les instances pédagogiques nationales ont intégré des cours classiques de fiabilité dans les différentes formations technologiques.

Le programme élaboré par le comité scientifique reflète l'orientation que nous avons voulu donner à cette manifestation. En plus des 05 Conférences plénières, 22 communications sélectionnées seront présentées en 04 sessions orales (et 01 session poster) : *« Sûreté de Fonctionnement et RBDO »*, « *Fiabilité dans les Sciences de l'Ingénieur* », « *Modèles Stochastiques de Fiabilité* » et « *Fiabilité dans les Réseaux de Télécommunication*».

Les Comités du Colloque tiennent à remercier tous ceux qui, de près ou de loin, ont apporter leur contribution à ce que cette 3ème édition (de la Conférence Nationale ***MFSI***) puisse tenir compte des évolutions scientifiques et technologiques de ces dernières années. La manifestation est d'ailleurs couplée avec la **6ème édition** de la « ***Journée Nationale de la Fiabilité*** », créée en 2013 par l'Ecole Nationale Polytechnique (El Harrach), l'AD – ENP (association des diplômés), en collaboration avec l'Unité de Recherche ***LaMOS*** Béjaia (Modélisation et Optimisation des Systèmes, http://www.lamos.org), à la mémoire du Professeur Abdelaziz Ouabdesselam (considéré comme étant le « père » des fiabilistes algériens).

A tous les participants à ***M.F.S.I.'2023***, nous souhaitons la bienvenue à Béjaia.

Le Président de la Conférence
Professeur Djamil Aïssani

Table des matières

I

CONFÉRENCES PLÉNIÈRES

Sommaire

Fiabilité des réseaux d'énergie électrique

Boudour Mohamed
USTHB (Alger)

Résumé— **En règle générale, l'évaluation de la fiabilité du réseau d'énergie électrique peut être divisée en deux parties, l'adéquation du système et la sécurité du système. L'adéquation du système concerne l'existence d'installations suffisantes (groupes de production) au sein du réseau pour satisfaire la demande ou la charge des consommateurs. Par conséquent, l'adéquation du système tient généralement compte de la situation statique, et non des perturbations auxquelles le réseau d'énergie est sujet. La sécurité ou la sureté de fonctionnement concerne la capacité du réseau à réagir aux perturbations dynamiques ou transitoires survenant dans le système. Ainsi, la sécurité prendra des mesures sur le système, quelles que soient les perturbations. En outre, le concept de fiabilité, qui est utilisé dans la plupart des études, concerne l'adéquation du système.**

Les méthodes d'évaluation de la fiabilité peuvent être divisées en deux familles : les méthodes déterministes et probabilistes. La plupart des approches publiées pour les grands réseaux d'alimentation électrique utilisent des méthodes probabilistes. Les méthodes probabilistes se concentrent sur une meilleure compréhension des impacts sur le fonctionnement du système associés à la nature intermittente notamment des sources d'énergie renouvelable.

L'évaluation de la fiabilité du système nécessite la simulation de son fonctionnement sur un temps long, malheureusement, hors de portée, vu la grande dimension du système et la complexité des phénomènes à simuler :

- **événements affectant le système (pertes d'ouvrages de production ou de transport, variations de la demande), avec les phénomènes transitoires qui s'ensuivent,**
- **actions des dispatchers (opérateurs du réseau) qui s'efforcent d'optimiser en permanence l'exploitation.**

Le concept LifeX : Perspectives et applications dans l'industrie pétrolière et gazière

Smati Abdelnacer
PEGAZ ENGINEERING
(Alger)

Résumé— **De nombreuses unités de l'infrastructure du secteur du pétrole et du gaz en Algérie approchent ou ont dépassé leur durée de vie originale de conception. Généralement, 20 à 25 ans sont considérés comme le délai de durée de vie sûre d'une installation. Pourtant, même après ce délai, de nombreux gisements pétroliers et gaziers produisent encore des niveaux substantiels d'hydrocarbures, en raison principalement des progrès dans les techniques de récupération, avec comme alternative soit le démontage et le remplacement de toutes les installations vieillissantes ou bien la prolongation de la durée de vie des installations existantes. L'industrie des hydrocarbures est fortement capitalistique et les investissements se comptent en milliards de dollars. C'est pourquoi il est important, d'un point de vue économique, d'envisager si certaines unités de cette infrastructure peuvent être utilisées ou réutilisées au-delà leur durée de vie prévue tout en tenant compte des considérations de fiabilité, sécurité et d'intégrité technique et opérationnelle. Le concept consistant à augmenter la durée de vie d'une installation sans augmenter les risques est appelé en terminologie anglo-saxonne Lifetime Extension ou LifeX. Ce concept assez récent a vu le jour vers au début des années 2000 et a depuis gagné en maturité et fait l'objet de nombreux travaux scientifiques. L'analyse LifeX nécessite de disposer de compétences et de connaissances diversifiées pour mener à bien un plan de prolongation de la durée de vie, en particulier :**

- **les mécanismes de vieillissement, les causes et les modes de défaillance,**
- **les modèles de dégradation dynamiques,**
- **Le choix des paramètres corrects pour décrire l'état actuel et futur des dommages,**
- **La modélisation de l'interaction de la maintenance sur les processus de dégradation.**

La fiabilité dans les sciences de l'ingénieur

Laggoune Radouane
Unité de Recherche LaMOS
Université de Béjaia

***Résumé*— La fiabilité est devenue une caractéristique incontournable dans la conception et l'exploitation des produits industriels, au point où elle est devenue un argument de vente adopté et systématiquement intégré par les fournisseurs dans les compagnes de promotion de leurs produits. Néanmoins pour garantir un certain niveau de fiabilité exigé par la spécificité de son produit, le concepteur devrait composer avec l'aléatoire dès le stade de conception. L'utilisateur du produit devrait ensuite composer également avec l'aléatoire pour maintenir ce niveau de fiabilité durant tout son cycle de vie ; en adoptant des stratégies d'inspection, de maintenance et de renouvellement adéquates et surtout optimales, notamment vis-à-vis des coûts. Tout cela n'est possible que par le recours à des modèles stochastiques de fiabilité. Durant cette conférence, nous tenterons de passer en revue un certain nombre de démarches d'application de modèles de fiabilité dans les sciences de l'ingénieur, elle s'agira notamment de l'optimisation des inspections et de la maintenance, de l'optimisation de la conception par la fiabilité (RBDO) ; dans les domaines de mécanique, de génie civil, de génie électrique ; dans divers secteurs tels que : les hydrocarbures, les réseaux de distribution de l'eau et de gaz, les ports, le textile,… Des exemples d'étude de cas seront également présentés.**

Fiabilité et estimation non paramétrique

Lagha Karima
Unité de Recherche LaMOS
Université de Béjaia

***Résumé*— La fiabilité est la discipline qui étudie les risques de défaillance d'un dispositif quelconque. Elle fait appel à la fois à des théories mathématiques, des connaissances d'ordre technologique et à l'expérience. L'un des principaux problèmes de la théorie de la fiabilité réside dans l'estimation des indices de fiabilité.**

La conférence a pour objectif de présenter différents résultats, de nos recherches sur l'estimation non paramétrique de la fonction fiabilité et la fonction taux de défaillance. On s'intéresse particulièrement à la méthode du noyau et la méthode des fonctions orthogonales.

Sécurité et fiabilité des systèmes informatiques

Hamza Lamia
Laboratoire d'Informatique
MEDicale (LIMED)
Université de Béjaia

Résumé— **Durant ces dernières décennies, la fiabilité et la sécurité sont devenues des caractéristiques incontournables de la conception et de l'exploitation des systèmes informatiques. La fiabilité d'un système informatique passe nécessairement par la prise en compte des éléments de sécurité indispensables, à savoir : la sécurité des matériels et des logiciels, la sécurité des droits d'accès, la sécurité des données et de leur sauvegarde. Le meilleur moyen d'assurer la sécurité et la fiabilité à la fois est d'identifier d'abord les problèmes potentiels avec les solutions et les coûts associés. L'ensemble des solutions doit être organisé sous forme d'une politique de sécurité cohérente à respecter par les utilisateurs du système à sécuriser. Néanmoins, l'apparition de nouvelles vulnérabilités engendre de nouveaux problèmes de sécurité dont la résolution représente un enjeu de taille pour la fiabilité. L'objectif de ce travail est de connaître l'influence de la sécurité sur la fiabilité des systèmes informatiques. Dans ce cadre, les techniques de sécurisation des systèmes informatiques seront présentées en vue d'assurer la fiabilité par la génération automatique de configurations de sécurité.**

II

FIABILITÉ DANS LES SCIENCES DE L'INGÉNIEUR

Sommaire

Évaluation et amélioration de la disponibilité des services Cloud data center par différentes technique de redondance Chez ICOSNET

Ait-Amra Ikhlef
Unité de Recherche LaMOS, Université de bejaia, Algérie
aitamaraikhlef@gmail.com

Outamazirt Assia
Unité de Recherche LaMOS, Université de bejaia, Algérie)
outamazirt.assia@gmail.com

Aïssani Djamil
Unité de Recherche LaMOS, Université de Bejaia, Algérie
djamil.aissani@univ-bejaia.dz

Morsli Ahmed
Icosnet, Centre d'Affaires El Qods, Chéraga, Algérie
Ahmed.morsli@icosnet.com

***Résumé*—**
Ce travail présente une évaluation de la disponibilité de l'infrastructure sous-jacente sur laquelle les services Cloud fournis par ICOSNET sont hébergés. Afin d'apprécier l'impact des différentes techniques de haute disponibilité sur la disponibilité des services Cloud, nous avons eu recours à la méthode de bloc diagrammes de fiabilité. Cette évaluation nous a permis d'identifier les points critiques susceptibles d'entraîner une réduction de la disponibilité des systèmes. En conséquence, des recommandations visant à améliorer la disponibilité des services Cloud ont été formulées pour chaque composant de l'infrastructure technique. Les résultats de cette étude revêtent une importance particulière pour le fournisseur de services Cloud ICOSNET, car ils lui permettront d'envisager des ajustements dans les stratégies de redondance et de haute disponibilité actuellement mises en œuvre au niveau de l'infrastructure sous-jacente du Cloud Computing.

***Mots-Clés*—**
Cloud Data Center d'ICOSNET, Disponibilité, Redondance, Haute Disponibilité, Blocs Diagramme de Fiabilité.

I. Introduction

Aujourd'hui, il existe plusieurs fournisseurs de service Cloud, à savoir publics, privés et hybrides [1]. Étant donné la diversité de ces fournisseurs, l'utilisateur du Cloud doit être capable de sélectionner celui qui répond au mieux à ses besoins. Du point de vue du fournisseur de service Cloud, de nombreux défis doivent être surmontés pour fournir des services qui répondent à toutes les exigences définies dans les accords de niveaux de service (SLA).

La haute disponibilité (HA ou High Availability) de l'infrastructure du fournisseur de service Cloud est la première chose à penser en terme de fourniture de service. Elle est représentée comme un mécanisme permettant d'assurer la continuité de la fourniture d'un service en qualité normale ou dégradée et, elle est étroitement liée à la gestion des défaillances des Data Centers. En effet, les pannes non planifiées des data centers sont coûteuses pour les fournisseurs et les utilisateurs des services Cloud en même temps. Une disponibilité élevée est atteinte lorsque le service en question est indisponible moins de 5.25 minutes par an, ce qui correspond à une disponibilité d'au moins 99.999% [2].

Cependant, garantir une disponibilité élévée du système d'information même en cas de sinistre, attaque ou panne informatique est un problème critique. Généralement, avant de valider un contrat SLA avec les clients, le fournisseur doit procéder à une évaluation de la disponibilité de l'infrastructure sur laquelle le service Cloud est hébergé. La plupart des fournisseurs offrent environ 99,999% de la disponibilité de service dans leur contrat SLA. Toutefois, les données réelles montrent que la valeur de la disponibilité du service Cloud pourrait être plus faible. Par conséquent, pour réduire les temps d'arrêt globaux dans les Data Centers et fournir une estimation fiable de la disponibilité du service, les fournisseurs doivent évaluer régulièrement leur infrastructure pour détecter les défaillance probables et réduire le temps nécessaire à la résolution de ces défaillance ou, encore, opter pour des mécanismes de haute disponibilité basées sur les techniques de redondance permettant une duplication de l'information pour assurer une disponibilité de service en permanence.

Dans cette étude, nous nous intéressons à évaluer la disponibilité de l'infrastructure sous-jacente sur laquelle les services Cloud, proposés par le fournisseur ICOSNET, un opérateur multiservices d'accès Internet, sont hébergés. L'objectif est d'analyser l'impact des divers mécanismes de haute disponibilité que nous introduirons au sein de chaque élément de l'infrastructure technique sur la disponibilité des données et des services Cloud. Pour ce faire, nous avons opté pour l'utilisation de la méthode des Diagrammes de Fiabilité par Blocs (DFB) afin de représenter nos systèmes de manière simplifiée et de les décomposer en blocs. Ces blocs constituent des composants disposés en série, en parallèle ou en configuration redondante k/n, ce qui facilite les calculs de disponibilité.

Le reste de cet article est structuté comme suit. Dans la Section II, nous présentons quelques concepts mathématiques de base nécessaires à une meilleure compréhension de la notion disponibilité et citons quelques techniques améliorant

la disponibilité dans les systèmes informatiques du Cloud Data Center. Dans la Section III, nous évaluons la disponiblité du système de stockage basé sur les RAIDs. La Section IV (resp. V) est dédiée à l'évaluation la disponibilité du Cluster Cloud d'ICOSNET (resp. réseau LAN) en proposant une nouvelle architecture du Cluster plus fiable capable de supporter une charge de deux défaillance de serveurs (resp. du LAN avec redondance). Dans la Section VI, nous introduisons l'analyse de la disponibilité de l'infrastructure technique du Cloud Data Center d'ICOSNET en utilisant plusieurs techniques de redondance. Enfin, la Section VII conclut l'article.

II. Aperçu

Pour bien mener notre étude et afin de mieux cerner et de répondre précisément à notre problématique, nous présentons à travers cette sections quelques concepts mathématiques de base nécessaires à une meilleure compréhension de la notion du disponibilité et citons des solutions de haute disponibilité actuellement utilisées par les fournisseurs de service Cloud dans le biais de l'informatique.

La disponibilité, $A(t)$, est définie comme étant la probabilité qu'un système soit opérationnel à chaque fois que son utilisation est requise [3] :

$$A(t) = \Pr(\text{Système est non défaillant à l'instant } t). \quad (1)$$

Elle constitue ainsi l'un des indicateurs clés de performance d'un système de stockage du cloud indiquant la fréquence à laquelle un système est en état de fonctionnement. Elle permet, en outre, de renseigner sur la rentabilité du système à la fois pour les fournisseurs de service, qui cherchent à éviter de payer des pénalités suite à une incapacité de satisfaire une exigence SLA, mais également pour les clients dont la rentabilité économique dépend de la continuité de leurs services. Le maintien de la disponiblité d'un équipement se résume à deux aspects : 1) s'assurer que l'équipement ne tombe pas en panne aussi longtemps que possible, 2) réparer l'équipement aussi rapidement que possible. L'indisponibilité

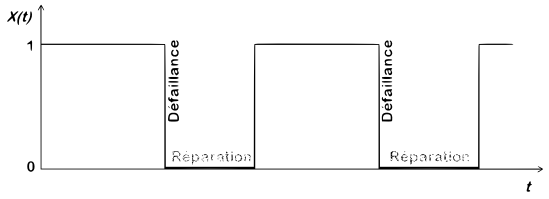


FIGURE 1. L'état du système

$\overline{A}(t)$, l'aptitude contraire, est la probabilité que le système soit défaillant à l'instant t.

La disponibilité moyenne, D, sur un intervalle de temps donné peut être évaluée par le rapport suivant :

$$D = \frac{MTBF}{MTBF + MTTR}, \quad \text{où,} \quad (2)$$

— MTBF (mean time between failures), le temps moyen entre pannes, est l'une des valeurs qui indique la fiabilité d'un système.

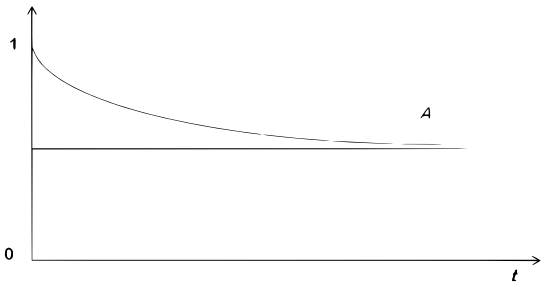


FIGURE 2. Disponibilité, indisponibilité

— MTTR (Mean Time To Repair), le temps moyen de réparation, est le temps moyen nécessaire pour réparer et restaurer un système défaillant.

A. Disponibilité des systèmes multi-composants

Un système se réfère généralement à toute entité formée d'un ensemble ordonné d'éléments indépendants et, il est conçu pour assurer une fonction bien déterminée. Cette dernière est définie par des relations que les composants du système entretiennent entre eux [4]. Le calcul de la disponibilité de ces composants nous permet de calculer la disponibilité de tout le système à l'aide de la méthode BDF [5].

Dans ce qui suit, nous citons quelques configurations des systèmes multi-composants dont on aura besoin dans la suite de ce travail. Pour chaque configuration, nous donnerons la formule mathématique qui nous permettera de calculer la disponiblité du système en question.

1) Système série: Un système série se caractérise par l'enchaînement linéaire de n composants (Fig. 3) et la défaillance de l'un de ses n composants entraîne la défaillance du système complet. La disponibilié de ce système, $A_S(t)$, est alors égale au produit des disponibilités de chaque composant :

$$A_S(t) = \prod_{i=1}^{n} A_i(t). \quad (3)$$

FIGURE 3. Système série

2) Système parallèle: Un système parallèle se caractérise par une association parallèle de tous les n composants (Fig. 4) et la défaillance de l'un ou de plusieurs composants est sans conséquence sur le système complet. La disponibilié de ce système est calculée comme suit :

$$A_S(t) = 1 - \prod_{i=1}^{n} (1 - A_i(t)). \quad (4)$$

3) Systèmes série-parallèle et parallèle-série: Un système série-parallèle (resp. parallèle-série) est constitué de n sous-systèmes connectés en parallèle (resp. en série) tel que chaque sous-système est composé de k éléments placés en série (resp.

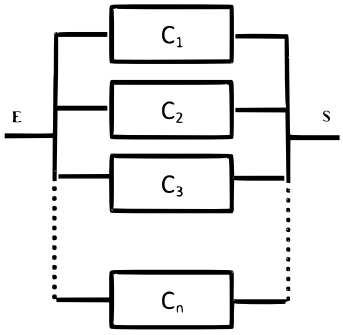


FIGURE 4. Système parallèle

en parallèle). La disponibilité du système série-parallèle (resp. parallèle-série) est calculée comme suit :

$$A_S(t) = 1 - \prod_{i=1}^{k}(1 - \prod_{j=1}^{n} A_{ij}(t)), \quad \text{où} \tag{5}$$

$$A_i(t) = \prod_{j=1}^{n} A_{ij}(t).$$

resp.

$$A_S(t) = \prod_{j=1}^{n}[1 - \prod_{i=1}^{k}(1 - A_{ij}(t))], \quad \text{où} \tag{6}$$

$$A_i(t) = 1 - \prod_{i=1}^{k}(1 - A_{ij}(t)).$$

4) Système redondant k/n: Un système redondant k parmi n fonctionne seulement si au moins k composants de ses n composants en parallèles fonctionnent (Fig. 5). La disponibilité de ce système est calaculée comme suit :

$$A_S(t) = \sum_{i=k}^{n} C_n^i A^i (1 - A)^{n-i}. \tag{7}$$

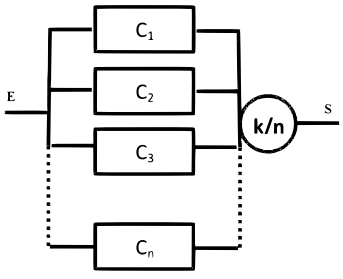


FIGURE 5. Système redondant k/n

B. Mesure du taux de disponibilité

La disponibilité se mesure souvent en pourcentage et elle est généralement décrite à travers le nombre de "9" (voir Tab. I) [6] :

TABLE I. Degrés de disponibilité d'un système

Disponibilité, D. en % et en Quant. de 9	*Qualification du système*	*Indiponibilité à l'année (min.)*	*Indisponibilité au mois (min.)*	*Signification pratique*
90% (un neuf)	D. non maitrisée	52596	4383	Indisponibilité de 5minutes/an
99% (deux neuf)	D. maitrisée	5259	438.3	Indisponibilité de 4 jours/an
99.9% (trois neuf)	D. bien maitrisée	525.9	43.83	Indisponiblité de 9 heures/an
99.99% (quatre neuf)	tolérant aux pannes	52.6	4.38	Indisponibilité de 1 heure/an
99.999% (cinq neuf)	hautement disponible	5.26	0.44	Indisponiblité de 5 minutes/an
99.9999% (six neuf)	très hautement disponible	0.53	0.04	Indisponibilité de 30 secondes/an
99.9999% (sept neuf)	ultra disponible	0.05	-	Indisponibilité de 3 secondes/an

C. Mécanismes de haute disponibilité

La haute disponibilité consiste à réduire les points de défaillances qui peuvent être détectés dans un système du cloud en mettant en place des actions et des paramètres techniques qui peuvent assurer la continuité de service en permanence. Cela est possible en appliquant certaines règles (par exemple la redondance des matériels et la mise en cluster, la sauvegarde (externalisation, centralisation sur site tiers), la réplication des données, la sécurisation des données (RAID),...) permettant la diminution de l'indisponibilité d'un système d'information.

III. MODÉLISATION DU SYSTÈME DE STOCKAGE RAID

Dans cette section et dans les sections qui suivent, nous nous intéressons à la modélisation et l'évaluation de la disponibilité, en utilisant la méthode des BDF, des systèmes Cloud exécutés sous l'infrastructure sous-jacente : stockage, serveurs et réseau, tout en considérant dans chaque système les mécanismes de haute disponibilité utilisés actuellement par l'opérateur ICOSNET. Ceci, dans le but de voir comment ces différents mécanismes contribuent à l'amélioration de la disponibilité en régime permanent des données et des services Cloud et d'identifier les points critiques qui diminuent cette dernière.

Les données que nous utilisons dans l'évaluation de la disponibilité des services Cloud proviennent de quatre sources [7]–[10]. Le tableau Tab. II détaille les valeurs du MTTF (temps moyen avant défaillance) et du MTTR pour chaque périphérique de l'infrastructure du Data Center que nous avons besoin dans notre étude. Le temps requis pour effectuer l'analyse de disponibilité est de 8760 heures (1an) et il a été calculé pour l'état stable. Les données sur les défaillances et les réparations de plusieurs composants sont modélisées et analysées à l'aide de distribution de probabilités et d'inférences statistiques. Les mesures opérationnelles MTTF et MTTR sont dérivées et utilisées pour estimer la disponibilité des composants.

Nous avons augmenté la valeur du MTTR pour les matériels réseau afin d'obtenir des résultats approximativement égaux à ceux du Data Center d'ICOSNET. Pour déterminer la disponibilité globale d'un système de stockage basé sur RAID, il est important d'évaluer avec précision la disponibilité du sous-système RAID. En utilisant la méthode des blocs diagramme de fiabilité, nous allons déterminer la probabilité

TABLE II. Valeurs MTTF, MTTR et la disponibilité des composants du Data Center

Composants	$\frac{1}{\lambda}$ *= MTTF (h)*	$\frac{1}{\mu}$ *= MTTR (h)*	*Diponibilité*
Alimentation serveur (Power)	$67 * 10^4$	1	0.999998507
CPU	$25 * 10^5$	1	0.9999996
NIC (Network Interface Cart)	$62 * 10^5$	1	0.999999839
RAM (Random Access memory)	$48 * 10^3$	1	0.999979167
VMM (Virtual Machine Monitor)	2893	2	0.999309154
VM (Virtual Machine)	2880	2	0.999306037
TOR Switch	$145 * 10^3$	8	0.99994483
Aggregation Switch	$2 * 10^5$	6	0.999970001
Router	$22 * 10^4$	6	0.999972728
Firewall	$14 * 10^4$	8	0.99994286
Link	19996	14	0.99930035

que les données sur différentes configurations RAID soient disponibles en régime permanent.

Afin de déterminer la différence de disponibilité entre les différents niveaux de configuration RAID, une valeur de disponibilité $D_{HDD} = 0.99952023$ est attribuée aux disques durs utilisés dans notre évaluation (où MTTF=200000h et MTTR= 96h), soit une indisponibilité de quatre jours dans une durée de 22 ans de fonctionnement, car les lecteurs de disque individuels sont moyennement fiables. Le MTTR est égal au temps de reconstruction du système (reconstituer le stockage plus restaurer les données).

A. Redondance par duplication

1) RAID-0 : Répartition de données (Striping): RAID-0 ne fournit aucune redondance des données. En d'autres termes, si un lecteur de l'ensemble RAID tombe en panne, toutes les données seront perdues. La Fig. 6 illustre le diagramme de fiabilité d'un ensemble de disques RAID-0. Les disques durs sont considérés comme étant en série.

FIGURE 6. Diagramme de fiabilité pour la structure RAID-0

En utilisant la Formule 3, la disponibilité et le temps d'arrêt moyen du système qui en résultent sont présentés dans le tableau Tab. III. Le RAID-0 n'est pas utilisé par les

TABLE III. Les résultats obtenus pour la structure RAID-0

Nombre de disques HDD	*2*
A_{RAID-0}	0.99904069
A_{RAID-0} (%)	99.90 %
Temps d'arrêt moyen (h)/an	8.40

fournisseurs Cloud car son fonctionnement ne permet pas la tolérance aux pannes et représente un risque important pour les données des clients.

2) RAID-1 : Mise en miroir (Mirroring): Le RAID-1 est le type de RAID utilisé par ICOSNET et il fournit une redondance complète des données.

Un disque dur peut échouer dans un ensemble couplé sans perte de données. Toutefois, si les deux lecteurs du même ensemble lié échouent, les données seront perdues. La Fig. 7 illustre le diagramme de fiabilité d'un ensemble de disques RAID-1. Les disques durs sont considérés comme étant en parallèle.

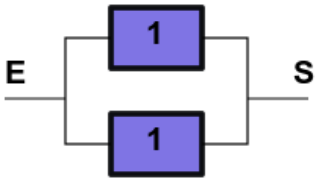


FIGURE 7. Diagramme de fiabilité pour la structure RAID-1

En utilisant la Formule 4, la disponibilité et le temps d'arrêt moyen du système qui en résultent sont présentés dans le tableau Tab. IV.

TABLE IV. Les résultats obtenus pour la structure RAID-1

Nombre de disques HDD	*2*
A_{RAID-1}	0.999999769
A_{RAID-1} (%)	99.99997 %
Temps d'arrêt moyen (sec)/an	7.28

D'après le résultat obtenu (l'indisponibilité de 7 seconde/an), nous constatons que le RAID-1 fournit une sécurité maximale des données en cas de panne d'un seul disque. Cette configuration est recommandée lorsque la sécurité est plus importante que la vitesse, comme le cas chez ICOSNET.

3) RAIDs Combinés : RAID-0+1 et RAID 10: **RAID-0+1 (Miroir de grappes)** : En RAID-0+1, les données sont agrégées par deux grappes sur un jeu de disques, puis mises en miroir sur un autre jeu de disques. Si un lecteur dans un jeu de disques échoue, ce jeu de disques est perdu, mais toutes les données resteront disponibles à partir du jeu de disques mis en miroir. Cependant, si l'un des disques durs du jeu de disques restant (le miroir) tombe en panne avant la restauration du premier jeu de disques, toutes les données sont perdues. La Fig. 8 illustre le diagramme de fiabilité d'un ensemble de disques RAID-0+1. Les disques durs sont considérés comme étant en série-parallèle.

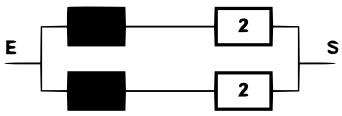


FIGURE 8. Diagramme de fiabilité pour la structure RAID-0+1

En utilisant la formule 5, la disponibilité et le temps d'arrêt moyen du système qui en résultent sont présentés dans le tableau Tab. V.

RAID 10 (Grappe de miroirs) : RAID 10 combine la mise en miroir de RAID-1 avec la répartition des données de RAID-0. Dans ce cas, les données sont réparties sur des ensembles

TABLE V. Les résultats obtenus pour la structure RAID-0+1

Nombre de disques HDD	***4***
$A_{RAID-0+1}$	0.999999079
$A_{RAID-0+1}$ (%)	99.99990 %
Temps d'arrêt moyen (sec)/an	29.04

de disques en miroir. Un seul disque dur dans un ensemble en miroir d'une matrice RAID 10 peut échouer sans aucune perte de données. Cependant, si les deux disques durs d'un ensemble en miroir échouent, toutes les données seront perdues. La Fig. 9 illustre le diagramme de fiabilité d'un ensemble de disques RAID-10. Les disques durs sont considérés comme étant en parallèle-série.

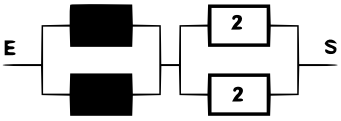


FIGURE 9. Diagramme de fiabilité pour la structure RAID-10

En utilisant le formule la 6, la disponibilité et le temps d'arrêt moyen du système qui en résultent sont présentés dans le tableau Tab. VI.

TABLE VI. Les résultats obtenus pour la structure RAID-10

Nombre de disques HDD	***4***
$A_{RAID-10}$	0.999999538
$A_{RAID-10}$ (%)	99.99995 %
Temps d'arrêt moyen (sec)/an	14.56

Les résultats obtenus montrent que les RAIDs combinés fournissent une indisponibilité supérieure à celle de RAID-1 (le RAID-0+1 : environ quatre fois supérieure ; RAID-10 : environ deux fois supérieure) dans une durée d'un an. C'est pour cette raison que le RAID-10 est préféré au RAID-01 dans la pratique.

B. Redondance par contrôle de parité

1) RAID-5: Dans RAID-5, les données sont réparties sur tous les disques durs. Lorsqu'un disque dur tombe en panne, toutes les données sont toujours disponibles. Les données manquantes sont recalculées à partir des disques durs restants et des informations de parité. La Fig. 10 illustre le diagramme de fiabilité d'un ensemble de disques RAID-5. Les disques durs sont considérés aussi comme étant un système redondant k/n.

Les résultats de disponibilité et d'indisponibilité obtenus pour les configurations RAID avec parité en fonction de n disques durs sont présentés dans le tableau Tab. VII et la Fig. 11.

D'après les résultats obtenus, nous avons constaté que plus le nombre de périphériques de stockage augmente, plus la disponibilité du système diminue. Cela est dû à la grande probabilité de subir des dommages matérielles avec un système

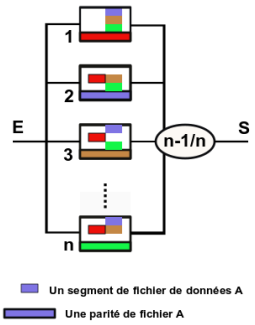


FIGURE 10. Diagramme de fiabilité pour la structure RAID-5

TABLE VII. Les résultats obtenus pour la structure RAID à parité unique

Nombre de disques HDD	***3***	***4***	***5***	***6***	***7***
$A_{RAID-3,4,5}$	0.99999930	0.999998619	0.99999770	0.999996551	0.999995173
$A_{RAID-3,4,5}$ (%)	99.99993 %	99.9998 %	99.9997 %	99.9996 %	99.9995 %
Temps d'arrêt moyen (sec)/an	22.07	43.55	72.53	108.76	152.22

comportant plusieurs lecteurs (6 par exemple) qu'un système à trois lecteurs (taux de défaillance augmente avec l'augmentation de nombre de disques).

En outre, nous pouvons constaté que les niveaux de RAID utilisant la mise en miroir risquent moins d'indisponibilité que ceux utilisant la parité.

Parmi les différentes configurations RAID avec parité, le RAID 5 est recommandé pour un bon équilibre entre protection des données et vitesse.

C. Redondance par double contrôle de parité

1) RAID-6: RAID-6 est similaire à RAID-5 mais les données de parité sont réparties en deux fois sur tous les disques durs. Ce mode de RAID peut prendre en charge une défaillance jusqu'à deux disques sans aucune perte de données. La Fig. 12 illustre le diagramme de fiabilité d'un ensemble de disques RAID-6. Les disques durs sont considérés aussi comme étant un système redondant k/n.

Les résultats de disponibilité et d'indisponibilité obtenus pour les configurations RAID avec double parité en fonction de n disques durs sont présentés dans le tableau Tab. VIII et la Fig. 13.

Les résultats indiquent que le RAID-6 offre une excellente disponibilité des données (100%) avec une indisponibilité nulle, même avec une configuration de huit disques durs.

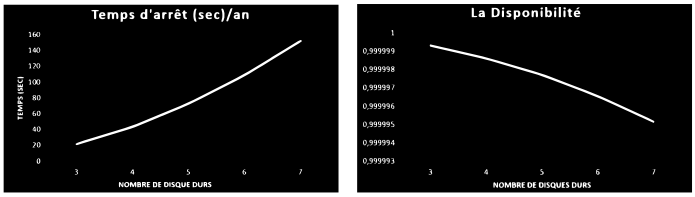


FIGURE 11. Présentation graphique des résultats

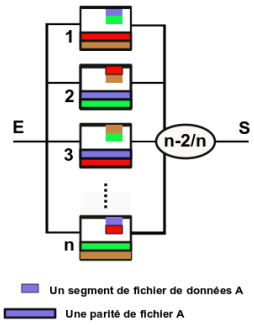


FIGURE 12. Diagramme de fiabilité pour la structure RAID-6

TABLE VIII. Les résultats obtenus pour la structure RAID à double parité

Nombre de disques HDD	*4*	*5*	*6*	*7*	*8*
A_{RAID-6}	0.999999999	0.999999998	0.999999997	0.999999996	0.999999993
A_{RAID-6} (%)	99.9999999%	99.9999998%	99.9999997 %	99.9999996%	99.9999993%
Temps d'arrêt moyen (sec)/an	0.031	0.063	0.094	0.126	0.220

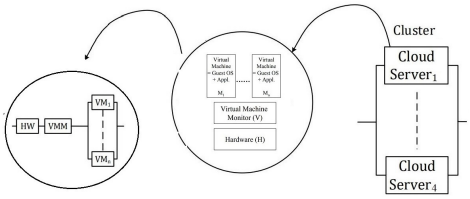


FIGURE 15. La configuration de la virtualisation dans un serveur Cloud

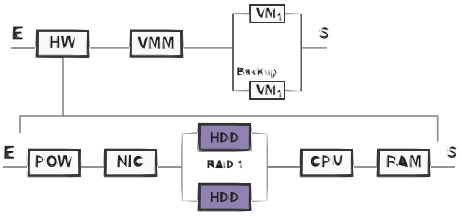


FIGURE 16. Digramme de fiabilité représentant le système serveur Cloud

La Fig. 14 résume les résultats obtenus sur la disponibilité et l'indisponibilité des systèmes RAID les plus couramment utilisés dans les entreprises.

IV. MODÉLISATION DU CLUSTER CLOUD

Nous examinons maintenant la disponibilité du Cluster Cloud d'ICOSNET. La Fig. 15 illustre la configuration de la virtualisation dans un serveur Cloud qui se trouve dans un système Cluster tolérant aux pannes (HA Cluster).

Chaque serveur est composé de : une unité centrale de traitement (CPU) ; des disques durs (HDDs) ; une carte d'interface réseau (NIC) ; une mémoire (RAM) ; une alimentation (POW) ; des machines virtuelles (VMs) ; un hyperviseur (VMM). La défaillance du CPU ou RAM ou HDD ou NIC ou POW ou VMM peut entraîner une défaillance de serveur. Ainsi, le diagramme de fiabilité du système serveur Cloud (système géré par le fournisseur) peut être représenté comme suit :

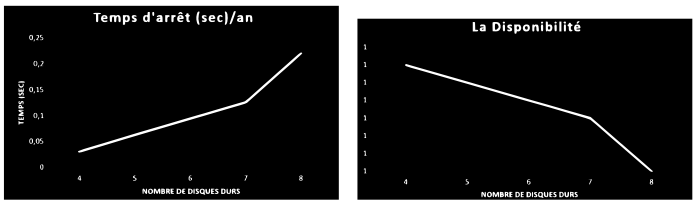


FIGURE 13. Présentation graphique des résultats

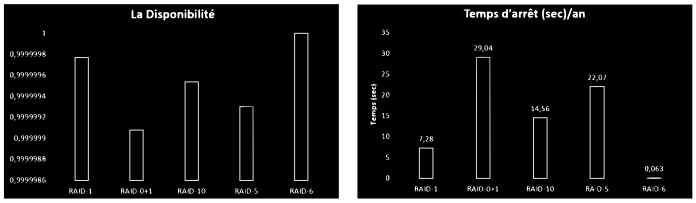


FIGURE 14. Diagrammes montrent les résultats obtenus pour les différents types de RAID

Les composants du serveur sont considérés comme étant un système en série. Nous avons présenté la VM et les HDD par des systèmes en parallèle, car les VMs sont redondées par réplication synchrone sur le sauvegarde (backup). Donc si une VM tombe en panne sur un serveur, son replica démarre sur l'autre hôte ; et le système de stockage utilise la configuration RAID-1.

Le calcul de la disponibilité du serveur Cloud nous a permis d'obtenir les résultats suivants (voir Tab. IX) :

TABLE IX. Les résultats obtenus pour le serveur Cloud

Composant du serveur	*Disponibilité*
Hardware	0.999976882
VMM	0.999309154
VM avec redondance	0.999999518
Cloud serveur	**0.99928557**

Comme on peut le voir dans le tableau Tab. IX, la disponibilité du serveur Cloud est : 99.928557 %. Ceci est inférieur à la disponibilité de chacun de ses composants. Il est clair que le composant qui a plus d'impact sur la disponibilité du serveur est le VMM, car sa disponibilité est très faible par rapport à d'autres composants, mais les serveurs placés dans le cluster ne sont pas autonomes (isolés). Dans un cluster, si un hyperviseur (VMM) tombe en panne, les machines virtuelles se déplacent automatiquement dans un autre hyperviseur de l'autre serveur qui appartenant à le même cluster. De même, si un serveur tombe en panne alors qu'il est en train de traiter des requêtes, d'autres serveurs du cluster prennent automatiquement la relève d'une manière aussi transparente que possible.

A. Modélisation du Cluster d'ICOSNET

Le HA cluster d'ICOSNET dispose de quatre serveurs et chaque serveur est relié par deux liens (Link) à deux commutateurs (Switchs). Le cluster peut prendre en charge une seule

défaillance de serveur sans interruption de service. La Fig. 17 illustre le diagramme de fiabilité du HA Cluster d'ICOSNET. Ce dernier peut être considéré comme un système redondant k/n.

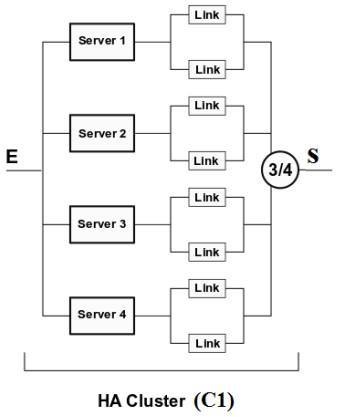


FIGURE 17. Digramme de fiabilité pour la configuration HA Cluster d'ICOSNET

Le calcul de la disponibilité du HA Cluster d'ICOSNET nous a permis d'obtenir les résultats suivants (voir Tab. X).

TABLE X. Les résultats obtenus pour la configuration HA Cluster d'ICOSNET (C1)

HA Cluster	*C1*
Disponibilité	**0.9999969362**
Temps d'arrêt moyen (sec)/an	96.61

Les résultats du cluster sont satisfaisants, d'où la disponibilité du service Cloud est améliorée. Cela est principalement dû aux mécanismes de redondance et de basculement des serveurs (mise en place d'un HA Cluster).

B. Nouvelle configuration du HA Cluster d'ICOSNET suggérée

Les résultats présentés dans la sous-section précédente peuvent encore être améliorés en proposant une configuration, C2, (voir la Fig. 18) de quorum évolutive plus fiable capable de supporter une charge de deux défaillance de serveurs (la moitié des noeuds) en même temps. Elle peut atteindre plus de disponibilité (environ moins trente secondes d'indisponibilité par an en moyenne) par rapport à la première configuration C1.

Le résultat de la disponibilité du HA Cluster avec la configuration (C2) est hautement satisfaisant (voir Tab. XI). On peut dire qu'elle offre une disponibilité de 100% et une indisponibilité nulle de service Cloud. Il s'agit là d'une amélioration majeure et cela est dû au mécanisme de quorum.

V. MODÉLISATION DU RÉSEAU LAN

Dans cette section, nous examinons la disponibilité du réseau LAN d'ICOSNET. La Fig. 19 illustre le diagramme de fiabilité de l'infrastructure réseau du Data Center d'ICOSNET.

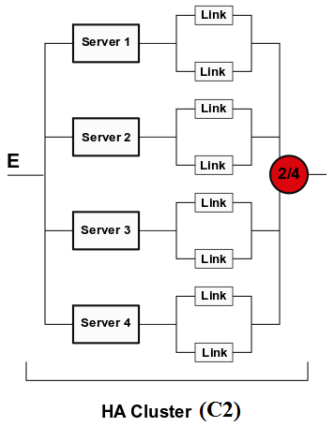


FIGURE 18. Diagramme de fiabilité pour la configuration HA Cluster d'ICOSNET suggérée

TABLE XI. Les résultats obtenus pour la configuration HA Cluster d'ICOSNET (C2)

HA Cluster	*C2*
Disponibilité	**0.9999999985**
Temps d'arrêt moyen (sec)/an	0.047

En décomposant le système LAN en série de sous-systèmes en parallèles et en série, la disponibilité de l'ensemble du système peut être facilement déterminée. Les résultats sont présentées dans le Tableau ci-dessous (voir Tab. XII) :

TABLE XII. Les résultats obtenus pour l'infrastructure réseau du Data Center d'ICOSNET

Composants	*Disponibilité*	*Temps d'arrêt moyen (min)/an*
Access Layer	0.999999996	0.0021
Aggregation Layer 1	0.999999467	0.280
Aggregation Layer 2	0.99994237	30.29
Core Layer	0.999999999	0.0005
LAN (A1)	0.999941832	30.57
LAN (A1)+ HA Cluster (C1)	**0.999938768**	**32.18**
LAN (A1)+ HA Cluster (C2)	0.999941830	30.57

Nous avons également effectué une analyse de disponibilité sur les différentes couches de réseau LAN ainsi que les deux configurations HA Cluster (C1 et C2) afin de déterminer les paramètres qui affectent de plus sur la disponibilité globale du l'infrastructure réseau. Les résultats indiquent que la couche d'agrégation 2 où il y a le pare-feu (Firewall) a le plus grand impact sur la disponibilité du l'infrastructure réseau, car sa

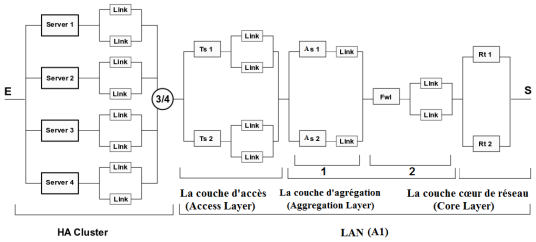


FIGURE 19. Diagramme de fiabilité pour l'infrastructure réseau du Data Center d'ICOSNET

disponibilité est très faible si on le compare avec d'autres composants. La panne de ce composant (Firewall) entraînera la perte de service.

Une disponibilité de degré quatre (9) n'est pas satisfaisante, mais elle peut être considérablement améliorée grâce au principe de redondance de tous les éléments critiques.

A. Nouvelle configuration LAN suggérée

Dans cette sous-section, nous proposons une nouvelle architecture LAN (A2) avec redondance dans les périphériques pare-feu (Firewall) et liens (Links). La Fig. 20 illustre le diagramme de fiabilité de l'architecture réseau LAN (A2) et les résultats obtenus sont présentés dans le tableau Tab. XIII.

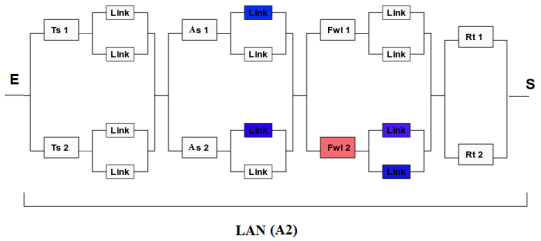


FIGURE 20. Diagramme de fiabilité pour LAN (A2) suggérée

TABLE XIII. Les résultats obtenus pour l'infrastructure réseau du Data Center avec la nouvelle configuration A2

Composants	*Disponibilité*	*Temps d'arrêt moyen (sec)/an*
Access Layer	0.999999996	0.1261
Aggregation Layer 1	0.999999999	0.0005
Aggregation Layer 2	0.999999996	0.1261
Core Layer	0.999999999	0.0005
LAN (A2)	0.99999999	0.3153
LAN (A2)+ HA Cluster (C1)	**0.999996926**	**96.93**
LAN (A2)+ HA Cluster (C2)	0.999999988	0.3626

La disponibilité qu'elle donne la deuxième architecture du LAN est différnete de celle donnée par la première architecture : une disponibilité presque de 100% du service Cloud surtout quand la configuration A2 suggérée combinée avec la configuration suggérée du HA Cluster (C2) et cela est dû au mécanisme de redondance des matériels d'accès au réseau.

A partir de cette évaluation sur le réseau LAN, nous avons pu répondre à ces deux questions suivantes : Comment décider quel composant devrait être répliqué ? Comment améliorer la disponibilité du LAN tout en réduisant les coûts ? Notre réponse est donc de rendre redondant les points critiques individueles disponibles.

L'indisponibilité de 96s/an est dûe à cause de la configuration HA Cluster (C1) (elle est quand même satisfaisante). Il est fortement recommandé de faire combiné LAN (A2) avec HA Cluster (C2) ou avec HA Cluster (C1) pour garantir une disponibilité maximale de service Cloud et une sécurité extrême des données de clients.

VI. APPLICATION SUR L'ARCHITECTURE GLOBALE DU DATA CENTER

Nous examinons maintenant la disponibilité de l'architecture globale du Data Center. L'entreprise ICOSNET possède un Data Center de Tiers III qui offre une disponibilité de 99.982%. La disponibilité de la fibre optique WAN est estimé avec une disponibilité de 99.961% (environ de 3h 30 min d'indisponibilité par an) par l'ingénieur d'infrastructure Cloud à l'entreprise ICOSNET en fonction de son expérience. La Fig. 21 illustre le diagramme de fiabilité de l'architecture globale du Data Center. Le système est considéré comme étant en série.

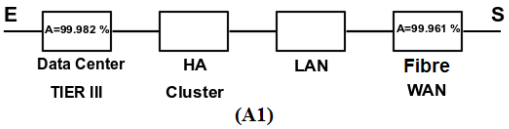


FIGURE 21. Diagramme de fiabilité pour l'architecture globale du Data Center d'ICOSNET

Nous évaluons la disponibilité globale (A1) en fonction des configurations du HA Cluster et du LAN. Les résultats sont présentées dans le Tableau ci-dessous (voir Tab. XIV) :

TABLE XIV. Les résultats obtenus pour l'architecture globale du Data Center

	HA Cluster(C1)+ LAN (A1)	*HA Cluster(C2)+ LAN (A2)*	*HA Cluster(C1)+ LAN (A2)*	*HA Cluster(C2)+ LAN (A1)*
Disponibilité (A1)	0.999368873	0.999430058	0.999426998	0.999371933
Temps d'arrêt moyen (h)/an	5.52	4.99	5.01	5.50

La disponibilité globale du Data Center est autour de trois (9) (environ de 5 heures d'indisponibilité par an), et nous n'avons observé aucune amélioration significative avec les deux architectures suggéréss HA Cluster (C2) et LAN (A2). Dans ce cas, on peut dire que la disponibilité du système dépend entièrement d'un/des point(s) critique(s).

Larchitecture globale du Data Center d'ICOSNET peut, malheureusement, y aller jusqu'à 5 heures d'indisponibilité de service par an. Cela est dû notamment à la fibre WAN qui n'est pas redondée par un deuxième point d'adduction et qui rencontre souvent des problèmes de pannes durant l'année.

A. Nouvelle confihuration du Data Center d'ICOSNET suggérée

Nous suggérons deux architectures (A2) et (A3) qui permettent d'améliorer considérablement la disponibilité du système globale en régime permanent, mais elles sont très couteuses financièrement et complexes à mettre en place. Notre objectif est de voir la variation de la disponibilité en utilisant les deux mécanismes de haute disponibilité : 1) la redondance des matériels d'accès au réseau et 2) la réplication sur un autre site distant. La Fig. 22 et la Fig. 23 illustrent les digrammes de fiabilité de l'architecture (A2) avec redondance dans la fibre WAN et l'architecture (A3) avec redondance d'un deuxième Data Center sur un autre site distant.

Nous évaluons la disponibilité globale de (A2) en fonction des configurations HA Cluster et LAN. Les résultats obtenus sont présentés dans le tableau Tab. XV.

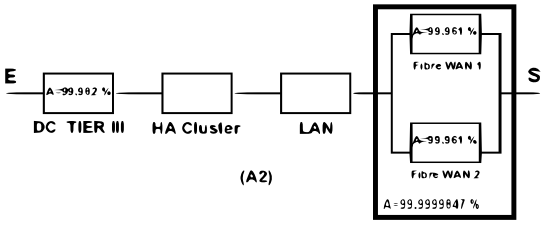


FIGURE 22. Diagramme de fiabilité pour l'architecture du Data Center (A2) suggérée

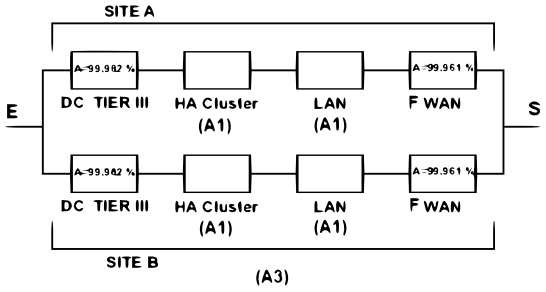


FIGURE 23. Diagramme de fiabilité pour l'architecture du Data Center (A3) suggérée

Et les résultats obtenus pour l'archetecture A3 sont présentés dans le tableau Tab. XVI.

Les résultats obtenus indiquent une amélioration importante de la disponibilité de service Cloud dans l'darchitecture (A2) par rapport à celle de (A1) (deux fois de plus supérieur). Cela est dû au mécanisme de redondance des matériels d'accès au réseau externe (la fibre WAN).

Pour l'archetecture (A3), on ignore le temps moyen de basculement d'un Data Center à un autre, la disponibilité de service Cloud est atteinte un niveau maximal (disponibilité de six (9)). Cela est principalement dû au mécanisme de redondance du Data Center sur un autre site distant (si le premier Data Center tombe en panne, son réplica démarre sur l'autre site et les clients accèdent aux mêmes données).

Avec la mise en place d'une reprise après sinistre du Data Center, la probabilité que les deux systèmes (deux sites) soient indisponible au même moment est très faible (P= $3.98*10^{-7}$).

TABLE XV. Les résultats obtenus pour l'architecture globale du Data Center (A2)

	HA Cluster(C1)+ LAN (A1)	*HA Cluster(C2)+ LAN (A2)*	*HA Cluster(C1)+ LAN (A2)*	*HA Cluster(C2)+ LAN (A1)*
Disponibilité (A2)	0.999758627	0.999819835	0.999816773	0.999761687
Temps d'arrêt moyen (h)/an	2.11	1.57	1.60	2.08

TABLE XVI. Les résultats obtenus pour l'architecture globale du Data Center (A3)

	Disponibilité	*Temps d'arrêt moyen (sec)/an*
A3	**0.999999601**	12.58

VII. CONCLUSION

La violation du contrat SLA et l'indisponibilité des services peuvent être converties en coûts directs pour le fournisseur de service Cloud ICOSNET. Pour cette raison, l'estimation de la disponibilité de l'infrastructure sous-jacente du Cloud peut aider le fournisseur de service Cloud ICOSNET à minimiser ses coûts. Dans ce travail, en utilisant des modèles basés sur les BDF, nous avons évalué la disponibilité du service exécuté sur l'infrastructure sous-jacente du Cloud et avons évalué l'impact de différentes stratégies de redondance et de tolérance aux pannes afin d'augmenter la disponibilité des services. Pour ce faire, nous avons d'abord effectué une application sur les systèmes de stockage RAID y incluant celui qui est utilisé actuellement par ICOSNET (RAID-1) afin de voir la différence entre eux en terme de disponibilité des données. Cela, nous a permis de constater que le RAID-6 est le meilleur choix pour la sécurité extrême des données.

Ensuite, nous avons effectué une application sur le Cluster Cloud. Les résultats obtenus montrent que la disponibilité du service est suffisante. La configuration du Cluster Cloud montre que ce dernier peut prendre en charge une seule défaillance de serveur sans interruption de service. Pour cette raison, nous avons proposé une autre configuration plus fiable capable de supporter une charge de deux défaillance de serveurs en même temps.

Enfin, nous avons effectué une application sur le réseau interne et l'infrastructure globale du Data Center. Les résultats obtenus montrent que la disponibilité des systèmes peut être améliorée de manière efficace et substantielle en se concentrant sur les composants qui ont plus d'impact sur la disponibilité (les points critiques individuelles disponibles).

Cette étude permettra à l'opérateur ICOSNET de voir approximativement l'état actuel de ses systèmes informatiques exécutés sous l'infrastructure sous-jacente du Cloud (stockage, serveurs, réseau) ainsi que leurs taux de disponibilité, de savoir comment les solutions de haute disponibilité contribuent dans l'augmentation du taux de disponibilité des systèmes informatiques et de modifier les stratégies de redondance adoptées actuellement au niveau de son Cloud Data Center.

A l'issue de l'étude effectuée, nous envisageons d'évaluer les coûts liés aux solutions de haute disponibilité mises en avant et de proposer un problème de minimisation de la consommation d'énergie dans le Cloud Data Center D'ICOSNET.

RÉFÉRENCES

[1] P. Mell and T. Grance "The NIST Definition of Cloud Computing," Special Publication 800-145, 2011. [online] https ://www.nist.gov/news-events/news/2011/10/final-version-nist-cloud-computing-definition-published.

[2] T. Maria and T. Francis, "*Service Availability : Principles and Practice*," A John Wiley & Sons, 2012.

[3] E. Elsayed, "*Reliability Engineering*," 2ème Édition. Addison Wesley Longman, Inc., 1996.

[4] R. Kaoutar, "*Optimisation de la disponibilité des systèmes multi-États*," Projet industriel de fin d'études, Université Moulay Ismail, 2015.

[5] M. Gilbert, "*Méthodologies appliquées : Adaptation au monde des réseaux*," 1er Édition. Ebook, 2017.

[6] E. Baucr and R. Adams, "*Reliability and availability of Cloud Computing*," A John Wiley & Sons-IEEE Press, 2012.

[7] F. Stenio, T. Eduardo, S. Marcelo, L. Victor and M. Paulo, "*Dependability assessment of virtualized networks*," In Communications (ICC), 2012 IEEE International Conference on, 2012.

[8] R. Souza, S. Marcelo and S. Fernandes, "*Importance Measures for NFV Data Center : An Availability Evaluation*," Workshop Pré-IETF (WPIETF'2018), 2018.

[9] J. Hlavacek and R Bestak, "Availability Model for Virtualized Platforms," *Advances in Electrical and Electronic Engineering*, 2013.

[10] H. Sandra, A. Xueli, K. Wolfgang, B. Sergio and K. Andreas, "*Data Center architecture impacts on virtualized network functions service chain embedding with high availability requirements*," In Proceedings of the IEEE Global Communications Conference, 2015.

Analyse fiabiliste appliquée à l'évaluation des performances des poutres à base de matériaux à gradient fonctionnel chargées en flexion

Allala Baya
Civil Engineering Department, University Mouloud Mammeri of Tizi Ouzou & Laboratory of Construction Engineering and Architecture (LGCA), Faculty of Technology, University of Bejaia, 06000 Bejaia, Algeria
baya.allala@ummto.dz

Si Salem Abdelmadjid
Civil Engineering Department, University Mouloud Mammeri of Tizi Ouzou & Laboratory of Construction Engineering and Architecture (LGCA), Faculty of Technology, University of Bejaia, 06000 Bejaia, Algeria
abdelmadjid.sisalem@univ-bejaia.dz

Ait Taleb Souad
Laboratory L2MSGC, University Mouloud Mammeri of Tizi Ouzou, Tizi Ouzou 15000, Algeria
souad.ait-taleb@ummto.dz

***Résumé*—**
Une analyse fiabiliste appliquée à l'évaluation des performances mécaniques des poutres à base de matériaux à gradient fonctionnel (FGM) en fonction du chargement extérieur appliqué, est menée dans le cadre de ce travail. La nouvelle conception des poutres proposée, consiste à associer deux matrices métallique et céramique, permettant d'améliorer la rigidité de la poutre, de s'opposer au développement des déformations latérales et enfin assurer la continuité entre la zone tendue et la zone comprimée de l'élément poutre. L'approche menée a pour but d'optimiser cette conception et d'évaluer l'influence de la variabilité des différents paramètres pour la détermination des états limites de défaillance des poutres composites, comparativement aux poutres classiques. La probabilité de défaillance ainsi que l'indice de fiabilité du système mécanique sont estimés par la méthode géométrique d'approximation du premier ordre FORM.

Mots-Clés—Poutres FGM, fiabilité, conception, probabilité de défaillance, performance.

I. INTRODUCTION

Actuellement, l'utilisation des matériaux composites a connu un essor important que soit dans l'industrie mécanique, dans l'aéronautique et surtout dans la construction civile. Ces matériaux sont susceptibles de substituer l'acier, en vu de leurs avantages qu'ils présentent à savoir : la résistance, la légèreté et l'insensibilité à la corrosion [1]. Les matériaux composites sont une solution très attractive pour répondre au besoin de renforcement et de réhabilitation des constructions en béton armé. L'amélioration des performances mécaniques des éléments de constructions en béton par utilisation de ces matériaux a été démontrée expérimentalement par de nombreux auteurs [2-4]. L'une des méthodes les plus utilisées est l'enroulement ou le collage d'une enveloppe composite sur les faces susceptibles à la rupture. Ces travaux ont confirmé que le mode de rupture, ainsi que la capacité portante des poutres renforcées par collage de plaques composites dépendent principalement des propriétés de l'interface béton-composite. Afin d'assurer la sécurité et le prolongement de la durée de vie de ces matériaux, qui sont limités par la dispersion de leurs propriétés mécaniques [5], notamment au niveau de leur résistance à la rupture, des méthodes fiabilistes sont développées et proposées par plusieurs chercheurs [6-8], permettant d'évaluer quantitativement le risque de défaillance compte tenu des incertitudes mises en jeu et également d'optimiser la conception pour des risques considérés.

L'approche fiabiliste repose typiquement sur les probabilités ou les distributions de probabilité des propriétés des matériaux et consiste à propager des incertitudes et/ou des méconnaissances dans des modèles de calcul spécifiques, afin d'obtenir une probabilité de défaillance d'une fonction du système étudié, pour un scénario de défaillance donné. Le concept fondamental de l'analyse de fiabilité est que les facteurs de résistance (R) et de chargement (S) sont des quantités statistiques ayant une tendance centrale (moyenne) et une dispersion autour de cette moyenne (écart-type). Les résistances et les sollicitations sont modélisées par des lois de probabilité intégrées dans la formulation des états-limites de rupture au moyen de la méthode d'analyse 'FORM' (First Order Reliability Method) basée sur la détermination de l'indice de fiabilité 'β', à travers une approximation géométrique ou numérique de la fonction d'état limite.

Dans ce travail, une analyse fiabiliste est menée sur des poutres FGM de dimension normalisée 160 x 80 x 1100 mm^3 chargées en flexion 04-points jusqu'à la rupture, pour mettre en exergue l'influence de la

variabilité des différents paramètres liés à la géométrie, aux propriétés mécaniques, et au chargement, pour la détermination des états limites de défaillance, ainsi que leur influence directe sur les modes de rupture des spécimens. Les résultats obtenus en termes d'évolution de l'indice de fiabilité ainsi que la probabilité de défaillance de la poutre en fonction du chargement extérieur appliqué sont présentés et confrontés avec ceux des poutres classiques avec une conception traditionnelle, dans le but de quantifier l'apport de cette technique de conception proposée en termes de performances mécaniques.

II. DESIGN ET OBJECTIF

En partant d'une approche d'actualité, basée sur la logique de développer de nouveaux éléments structuraux, qui présentent à la fois un poids propre réduit et des performances mécaniques élevées, une nouvelle technique de conception des poutres composites est proposée. La conséquence directe de cette conception est une réduction significative des quantités des interactions, conduisant à un double bénéfice : réduction du rapport de masse de la structure, et amélioration de ces performances mécaniques, notamment la rigidité flexionnelle, la capacité portante et la ductilité. Aussi, assurer la continuité dans toute la section transversale de l'élément poutre à travers les différentes mailles qui constituent la poutre FGM (Voir fig. 1).

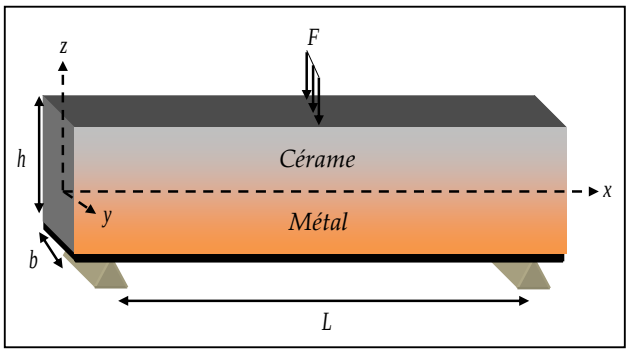


Figure 1. Conception de la poutre FGM étudiée.

III. ANALYSE FIABILISTE

La théorie de la fiabilité vise à déterminer la probabilité de défaillance d'un système de structures à l'aide d'une formulation aléatoire des données du problème et de la modélisation du mode de défaillance, à travers un couplage mécano-fiabiliste.
Cette probabilité peut être obtenue par intégration directe de la fonction de densité conjointe sur le domaine de défaillance. Elle est donnée par fig. 1.

$$P_f = P\left[G\left(\{X\}\right) \le 0\right] = \int_{D_f} f_{x_x}\left(x_i\right)\, dx_1 ... dx_n \qquad (1)$$

La valeur de la probabilité de défaillance, en principe est faible, le calcul analytique conduit à des erreurs importantes car Pf peut-être de l'ordre de grandeur de l'erreur d'intégration. Par conséquent les méthodes d'intégration directe ne peuvent être menées, que dans des cas particuliers favorables où les fonctions de densités ont des formes simples. Le calcul de Pf est alors envisagé par des approches d'approximation.

A. *Approche FORM*

La méthode F.O.R.M est basée sur la détermination de l'indice de fiabilité 'β', à travers une approximation géométrique et numérique de la fonction d'état limite. Elle nécessite de travailler dans un espace probabiliste réduit, c'est-à-dire un espace dans lequel, toutes les variables aléatoires sont transformées en lois normales centrées réduites et indépendantes.

L'indice de fiabilité 'β' est défini comme étant la distance minimale de l'origine de l'espace réduit à un point de la surface d'état limite. Une fois l'indice de fiabilité est déterminé, une probabilité de défaillance conventionnelle lui est associée. Elle s'exprime selon la fonction de répartition de loi normale standard en fonction des moyennes et des écart-types des fonctions résistance et sollicitation respectivement, comme indiqué par fig. 2.

$$P_f = \phi(-\beta) = -\phi\left(\frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}\right) \qquad (2)$$

B. *Couplage mécanique-fiabilité*

Le scénario de défaillance de la structure défini est lié à la rupture de la poutre par le moment fléchissant développé sous un chargement de flexion 4-points. Le critère de défaillance est défini à l'aide de la fonction d'état limite donné par fig. 3.

$$G\left(\{X\}\right) = M_R - M_S \qquad (3)$$

M_R : Moment résistant développé par la section (fonction de résistance « R »).

M_S : Moment sollicitant extérieur (fonction de sollicitation « S »).

La marge de sûreté de la poutre FGM s'annule lorsque l'ouverture des fissures dépasse la valeur limite définit par la fonction d'état limite de fig. 4. Quand l'état limite est atteint, la défaillance aura lieu lorsque le moment résistant équivaut au moment sollicitant.

C. *Modélisation des variables aléatoires*

Les variables aléatoires, concernant la fonction d'état limite de fig. 4 pour une poutre FGM sont de trois

catégories : les variables liées à la géométrie, celles liées aux propriétés mécaniques internes du système et celles liées au chargement. Elles sont modélisées selon une loi normale. Tab. 1 résume les paramètres de position et de dispersion des fonctions, de résistances et de sollicitations respectivement. Les lois de probabilité permettent, ainsi de décrire de manière théorique le caractère aléatoire des matériaux utilisés à travers des densités de probabilités.

Table 1. Modélisation des variables aléatoires

Paramètres	*Moyenne (KN)*	*Ecart-type (KN)*	*C.O.V (%)*	*Loi de distribution*
M_R	21.34		7.8 %	Normale
M_S	13.54		10 %	Normale

Les variables aléatoires de l'espace physique, sont transformées en des variables centrées réduites et indépendantes. Différentes transformations iso-probabilistes permettent ainsi de faire ce passage [9, 10]. La transformation utilisée dans le cadre de cette étude est celle de Rosenblatt [11] donnée par fig. 4 en fonction du vecteur des variables aléatoires Xi.

$$U_i = \frac{(X_i - \mu_i)}{\sigma_i} \tag{4}$$

Cette transformation permet de déterminer la distance minimale (β) de l'origine à la surface d'état limite H(U)=0 dans l'espace normé centré réduit. Cette distance définit un hyperplan tangent à la fonction d'état limite. Fig. 1 montre les densités des probabilités conjointes dans l'espace standard du vecteur des variables aléatoires M_S et M_R des fonctions sollicitation et résistance.

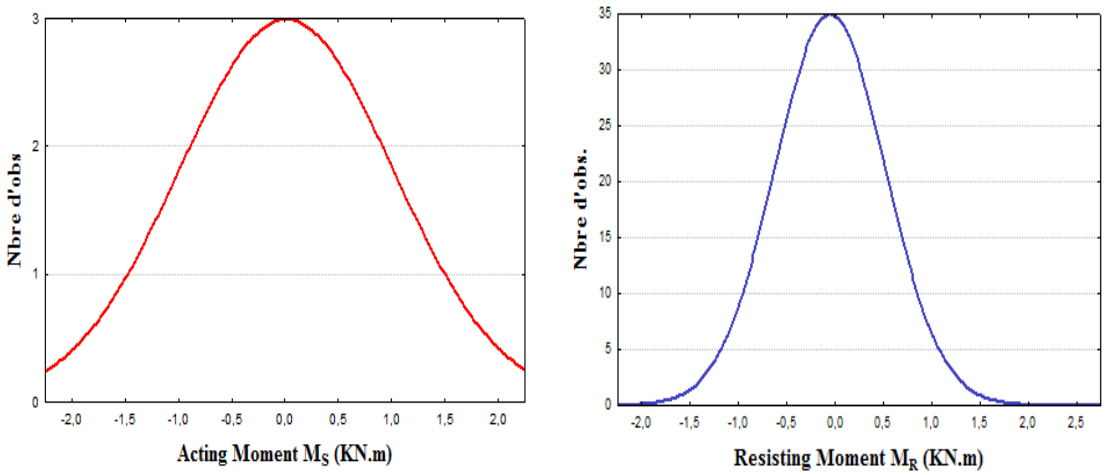


Figure 2. Densité des probabilités dans l'espace normé centré réduit.

IV. INTERPRETATION ET ANALYSE

Un modèle fiabiliste spécifique est développé, permettant de représenter l'évolution de l'indice de fiabilité ainsi que la probabilité de défaillance du système mécanique en fonction du chargement extérieur de flexion appliqué au moyen de la méthode FORM. Tab. 2 présente les résultats de l'analyse fiabiliste menée sur deux spécimens d'étude à savoir la poutre FGM et la poutre de référence en béton armé avec un ferraillage traditionnelle (2T6 placées dans la zone tendue en flexion) pour différente taux de chargement considéré. Fig. 3 montre l'évolution de la probabilité de défaillance en fonction de l'évolution de la capacité portante des structures considérées au cours du chargement.

La probabilité de défaillance due à l'amorce et la propagation des fissures de flexion dans le système - mécanique considéré, augmente significativement en fonction du chargement extérieur appliqué Cette probabilité est nulle au début du chargement pour les deux systèmes considérés, puis elle se propage jusqu'à la rupture totale des éléments. À première vue, on constate un sérieux décalage entre la probabilité de la poutre développée et celle de référence.

Table 2. Résultats de l'analyse fiabiliste

Chargement appliqué P (kN)	*Poutre FGM*		*Poutre en BA*	
	β	P_f	β	P_f
3	5.53	2.36E-7	5.46	6.65E-4
5	4.31	3.32E-5	3.35	0.09
10	2.87	0.26	1.85	0.91
12	2.35	0.64	0.07	0.94
15	2.11	0.85	0.00	1.00
20	1.61	0.98	0.00	1.00

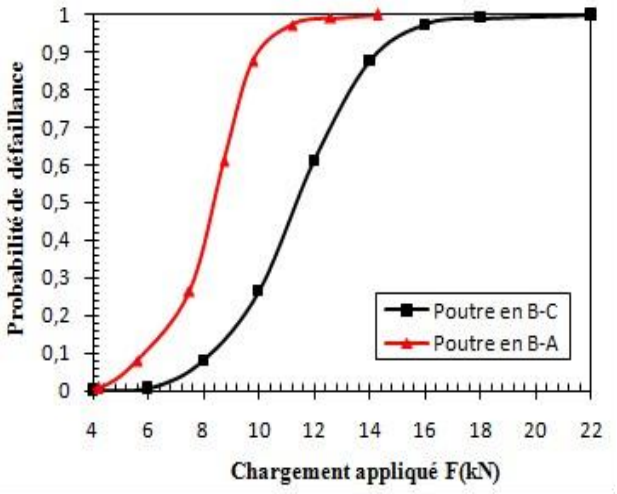


Figure 3. Evolution de la probabilité de défaillance

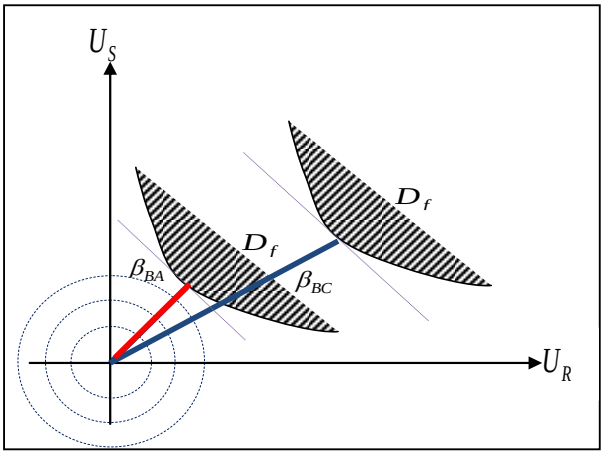


Figure 4. Illustration graphique de l'indice de Fiabilité de la méthode FORM

Un complément de résultats obtenus de l'analyse fiabiliste par la méthode FORM est illustré sur Fig. 4, qui représente les domaines de défaillance et les fonctions de performance, dans l'espace standard correspondante à un chargement extérieur appliqué de 10KN. Les courbes montrent clairement l'amélioration en termes de performances mécaniques et de capacité portante de la poutre FGM.

L'indice de fiabilité qui reflète le risque d'amorçage des fissures de flexion, diminue de façon significative en fonction du chargement extérieur appliqué (voir Tab. 2). Ceci est traduit par la propagation des fissures de flexion dans les éléments, jusqu'à l'état limite défini par le modèle fiabiliste. La fiabilité de la poutre de référence diminue rapidement, la rupture est atteinte pour une capacité portante maximale de 13,12 KN. La marge de sûreté de la poutre développée, s'annule pour une capacité portante de 22,32 KN.

L'indice de fiabilité de la poutre composite développée, correspondant à l'apparition de la première fissure est défini comme étant la distance minimale de l'origine de l'espace réduit au point de la surface d'état limite vaut $\beta = 2.37$. La confrontation montre que la poutre de référence présente un domaine de sureté moins inférieur à celui de la poutre FGM, l'indice de fiabilité de la poutre de référence, correspondant à l'apparition de la première fissure vaut $\beta = 1,83$. L'apport en termes de performances mécaniques est estimé à 60%. À titre d'illustration, la probabilité de défaillance de la poutre composite vaut environ 26 % pour F = 10 KN, alors la poutre de référence présente une probabilité de défaillance de l'ordre de 90% pour le même taux de chargement. Ces résultats sont en bonne corrélation avec ceux obtenus par [12-20]

V. CONCLUSION

Cette analyse met en évidence l'influence de la variabilité des paramètres des fonctions de chargement et de résistance sur la fonction d'état limite, à travers un modèle fiabiliste appliqué sur des poutres composites à base de matériaux à gradient fonctionnel, soumises à un chargement de flexion 04-points. Cette conception proposée offre une réduction de poids propre de 20% qui peut être transformée en charge d'exploitations additionnelles. Le gain en performances mécaniques (capacité portante) estimé est de l'ordre de 60%.

Le traitement des différents résultats obtenus en termes d'évolution de l'indice de fiabilité qui reflète l'état de sureté de l'élément, et en termes de probabilité de défaillance due à l'amorce et la propagation des fissures de flexion laisse entrevoir tout l'intérêt de l'utilisation des matériaux composites pour améliorer les performances mécaniques, notamment la capacité portante des poutres composites sous chargement de flexion.

La suite du travail portera sur l'optimisation de la conception, à travers une étude de sensibilité statistique de chaque variable aléatoire liée au système mécanique, ainsi que la modélisation des différents scenarios de défaillance susceptibles de se produire, par une approximation quadratique de la surface de défaillance.

REFERENCES

[1] J.M. Berthelot., " *Matériaux composites, comportement mécanique et analyse des structures* " , Edition. Lavoisier, 2005, ISBN: 2-7430-0771-0.

[2] Bentayeb F, Ait Tahar K., Chateauneuf A., " *New technique for reinforcement of concrete columns confined by embedded composite grid* ", Construction and Building Materials 22 (2008) 1624–1633.

[3] S. Djenad, et al. Finite element modeling of partially-confined concrete and RC columns with embedded hexagonal-FRP strips under axial and horizontal loading. Structures 54 (2023) 369–385

[4] G. Vasudevan and S. Kothandaraman: " *Study on Non-Linear Flexural Behavior of Reinforced Concrete Beams using Ansys By Discrete Reinforcement Modeling* " Strength of Materials, Vol. 45, No. 2, March, 2013.

[5] H. Dehmous, H. Welemane., " *Multi-scale reliability analysis of composite structures – Application to the Laroin footbridge* ". Engineering Failure Analysis 18 (2011) 988–998.

[6] S.C.Ribeiro, S.M.C. Diniz., " *Reliability-based design recommendations for FRP reinforced concrete beams* " Engineering Structures 52 (2013) 273–283.

[7] B. Behnam, C. Eamon., " *Reliability-based design optimization of concrete flexural members reinforced with ductile FRP bars* ", Construction and Building Materials 47 (2013) 942–950.

[8] PL. Liu, A. Der Kiureghian., " *Multivariate distribution models with pre described marginal and covariances* ". Probab Eng Mech 1986; 1(2).

[9] Frangopol DM, Ide Y, Spacone E, Iwaki I. " *A new look at reliability of reinforced concrete columns* ". J Struct Saf (1996); 2(18):123–50.

[10] H. Dehmous, H. Welemane, M. Karama, K. Ait tahar. "*Reliability approach for fibre-reinforced composite design* ", -International Journal for simulation and multidisciplinary design optimization, IJSMDO/2008/ISI/02, vol 2 p1-9 Ed. EDP.

[11] M. Lemaire., " *Fiabilité des structures: couplage mécano-fiabiliste statique* ", Edition. Hermès- Lavoisier, Paris, 2005.

[12] C. *Zhou et al. Reliability and sensitivity analysis of composite structures by an adaptive Kriging based approach. Composite Structures 278 (2021) 114682*

[13] *P.M. Zadeh and M. Mohaghegh. An efficient Bi-level hybrid multi-objective reliability-based design optimization of composite structures. Composite Structures 296 (2022) 115862*

[14] *H. Sharma, R. Ganguli Optimization of a higher-order sandwich composite beam under uncertainties. Composite Structures 269 (2021) 114003*

[15] *N. Li et al . Optimal design and strength reliability analysis of pressure shell with grid sandwich structure. Ocean Engineering 223 (2021) 108657*

[16] *A. Ameryan et al. Investigation of shear strength correlations and reliability assessments of sandwich structures by kriging method. Composite Structures 253 (2020) 112782*

[17] *G. Qi, et al. Modeling and reliability of insert in composite pyramidal lattice truss core sandwich panels. Composite Structures 221 (2019) 110888*

[18] A. SI SALEM, et al. *Experimental and statistical investigation of a new concrete-composite beam with encased polymer tube wrapped by FRP* ; Frontiers of Structural and Civil Engineering, 9(2): 154-162 (2015)

[19] S. Ait Taleb, et al. Experimental and theoretical modeling coupled to a reliability approach for flexural failure prediction in hybrid composite beams. Asian Journal of Civil Engineering (2020) 21:495–504

[20] S. Medjmadj, A. Si Salem and S. Ait Taleb. *Experimental behavior of plaster/cork functionally graded core sandwich panels with polymer skins*. Construction and Building Materials 344 (2022) 128257

Simplified equations of stress and strain for a shrink-fit assembly

Allal Bedlaoui
Laboratoire Energétique, Mécanique et Ingénieries, Université de Boumerdes, 35000 Boumerdes, Algeria
e-mail: a.bedlaoui@univ-boumerdes.dz,

Hamid Boutoutaou
Laboratoire Energétique, Mécanique et Ingénieries, Université de Boumerdes, 35000 Boumerdes, Algeria

Fadila Guerrache
Laboratoire Energétique, Mécanique et Ingénieries, Université de Boumerdes, 35000 Boumerdes, Algeria

Abstract –

Shrink-fit Assemblies have been used to produce a more robust running surface for wooden wheels for a long time. Generally, the two parts are cylindrical or conical. Shrink-fit is now a process that involves creating contact between two cylinders, there is no third party, and this is good economically. It's employed in a variety of industries, including automotive, aerospace, oil and gas, and train wheels. To perform this operation, the inner cylinder's outer radius must be greater than the outer cylinder's inner radius, that difference between them is called 'interference', the latter is being important in the assembly because it contributes to increasing the resistance of the assembly. There are three ways to do this: the first by heating the outer cylinder until it expands, the second by cooling the inner cylinder until it shrinks and the third way is to realize the fitting under a press. When two cylindrical components are assembled by pushing or shrinking one onto the other, a contact pressure and friction force is created at the interface between the two matching parts. In this work, We investigate a shrink-fit assembly made up of two hollow cylinders, one of which is exposed to internal pressure, we neglect the interference then we compare the simplified equations for radial, hoop and Von-Mises stress and strain with the results of numerical simulations using the finite element method. This is to demonstrate that simplified equations can be adopted in certain situations when the interference is very small.

Keywords—

Radial stress, *Hoop stress*, *Strain*, *Interference*, *Shrink-fit.*

I. INTRODUCTION

In these years, rapid developments in the domain of mechanical assemblies have made important progress in the development of fretted assemblies in terms of complex configurations. In the fields of use such as aerospace and transport. In a variety of mechanical parts, shrink-fit is a vital connection, the absence of a third party and this is what makes it economical. Despite its relevance, this contact problem has only a few three-dimensional finite element assessments. Because in general, evaluating contact between two or more bodies is a particularly difficult non-linear behavior. This type of assembly relies on interference to obtain coherence, but in theory, it is difficult to rely on these calculations, so it is better to rely on simplified equations in some cases when the interference is very small. Initially, we extract the simplified equations by neglecting the interference value, then we compare the simplified equations with the original and then with the simulation results. The assembly of two hollow cylinders has been the topic of several research. Zhang et al (1999) made a comparison between the analytical method based on Lame equation and the finite element method for assembling a component of two thick cylinders. The result showed that Lame equations have limitations since they consider perfect the contact surface[1,2].

N. laghzale and H. Bouzid (2016) compared the results of the analytical method with the results of the finite element method in shrink-fit assembly for several interferences values to calculate the maximum interference value to avoid damage to the shrink-fit assembly. A. Mouaa and al (2019) offer an analytical methodology for measuring the stresses in shrink-fitting an elastic solid shaft and an elastic-plastic cylindrical hub. In the elastic field, M. Belhou and N. Laghzale (2019) investigated the influence of interference on the distribution of residual stress for FGM. These three studies were conducted on a two

hollow cylinder assembly. U.A. Compos and D.E. Hall (2019) explained that Lame's equation can be simplified in measuring the hoop stress and contact pressure (radial stress) in two thin-walls cylinder assembly by neglecting the thickness. P. Pederson (2006) used two methods of distributing contact pressure without incrementation, the first determining the shape of a shrink-fit surface and the second using the super elements technique. The assembly used in this article consists of two hollow cylinders[3–7]. In this paper, we study the effect of simplified equations on a shrink-fit assembly consisting of two hollow cylinders, under which the inner cylinder is subjected to pressure. First, a comparison is made between the original and simple equations, then we compare the results of the simplified equations with the simulation results. The simplification of the equations for the radial, circumferential and Von Mises stresses lies in the neglect of the interference value, and the results showed a great agreement between the original and simplified equations.

II. THE SIMPLIFIED RADIAL, HOOP AND VON MISES STRESSES EQUATION

In this case, we are studying the plane problem of shrink-fit via two cylinders are assembled (Fig. 1). With the following parameter: the inner ring's inner radius (a), the inner ring's nominal outer radius equal to the outer ring's notional inner radius (b), the outer ring's outer radius (c), and the shrink-fit interference $e = b_i - b_0 \geq 0$ [7].

For the simplified radial, hoop, and von Mises stresses and strain equations, make the following assumptions:

1) Both component have the same material properties

2) The thickness of the two cylinders should be equal and thin $\frac{b}{c} < 1.5$.

3) Only linear elastic shrink-fits with no friction and no plastic strain are addressed.

4) The inner cylinder is under internal pressure p.

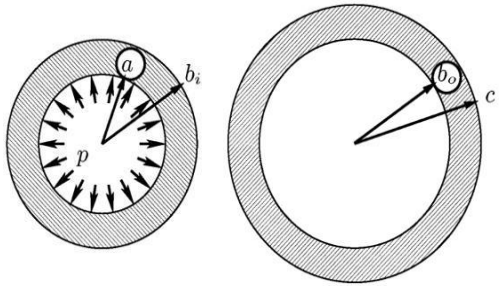


Figure 1. Internal pressure p and shrink fit interference e in a shrink fit model [7]

The radial, hoop, and Von Mises stresses functions for the cylinders are radial position r, internal pressure p, interference value e, and radii a, b, and c [7]:

For inner part $a \leq r \leq b$

$$\sigma_r = 1/2b(c^2 - a^2)\left(-p(2a^2b)(c^2/r^2 - 1) - eE(c^2 - b^2)(1 - a^2/r^2)\right) \quad (1)$$

$$\sigma_\theta = 1/2b(c^2 - a^2)\left(-p(2a^2b)(-c^2/r^2 - 1) - eE(c^2 - b^2)(1 + a^2/r^2)\right) \quad (2)$$

For outer part $b \leq r \leq c$

$$\sigma_r = 1/2b(c^2 - a^2)\left(-p(2a^2b)(c^2/r^2 - 1) - eE(b^2 - a^2)(c^2/r^2 - 1)\right) \quad (3)$$

$$\sigma_\theta = 1/2b(c^2 - a^2)\left(-p(2a^2b)(-c^2/r^2 - 1) - eE(b^2 - a^2)(-c^2/r^2 - 1)\right) \quad (4)$$

The Von Mises stress and strains in the absence of axial stresses $\sigma_z = 0$, are calculated as follows [3]:

$$\sigma_{VM} = \sqrt{\left(1/2(\sigma_r - \sigma_\theta)^2 + \sigma_r^2 + \sigma_\theta^2\right)} \quad (5)$$

$$\sigma_{VM} = E.\varepsilon \Rightarrow \varepsilon = \sigma_{VM}/E \quad (6)$$

In the elastic field, the interference value is very small so it can be neglected to calculate radial, hoop, Von Mises stress and strain. The resulting equation after simplification are:

For inner part $a \leq r \leq b$

$$\sigma_{rs} = 1/2b(c^2 - a^2)\left(-p(2a^2b)(c^2/r^2 - 1)\right) \quad (7)$$

$$\sigma_{\theta s} = 1/2b(c^2 - a^2)\left(-p(2a^2b)(-c^2/r^2 - 1)\right) \quad (8)$$

For outer part $b \leq r \leq c$

$$\sigma_{rs} = 1/2b(c^2 - a^2)\left(-p(2a^2b)(c^2/r^2 - 1)\right) \quad (9)$$

$$\sigma_{\theta s} = 1/2b(c^2 - a^2)\left(-p(2a^2b)(-c^2/r^2 - 1)\right) \quad (10)$$

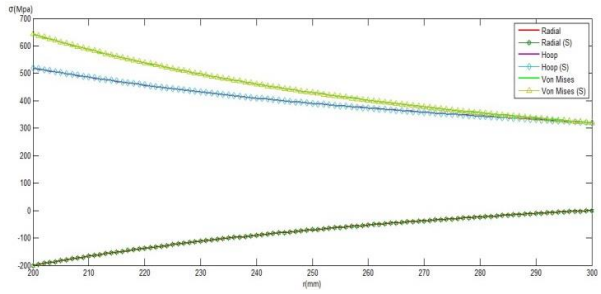
So here's the Von Mises stress equation and the strain: [3]:

$$\sigma_{VMs} = \sqrt{\left(1/2(\sigma_{rs} - \sigma_{\theta s})^2 + \sigma_{rs}^2 + \sigma_{\theta s}^2\right)} \quad (11)$$

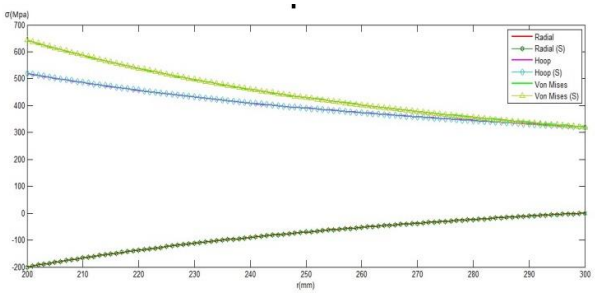
$$\sigma_{VMs} = E.\varepsilon_s \Rightarrow \varepsilon_s = \sigma_{VMs}/E \quad (12)$$

For example, consider the inner part and outer part of steel that has the following properties: E=2×10^5 Mpa, υ = 0.3, a=200mm, b=250mm, c=300mm, p=200Mpa.

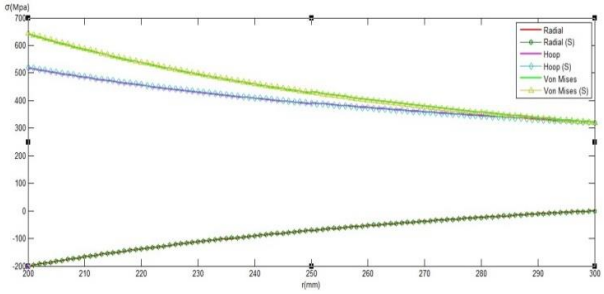
Fig.2 and Fig.3 shows that the difference between the curve data from radial, hoop, Von Mises stresses and strain Equations (1), (2), (3), (4), (5), (6) and simplified Equations (7), (8), (9), (10), (11), (12) respectively for different value the interference, since it is clear that there is a difference but is negligible.



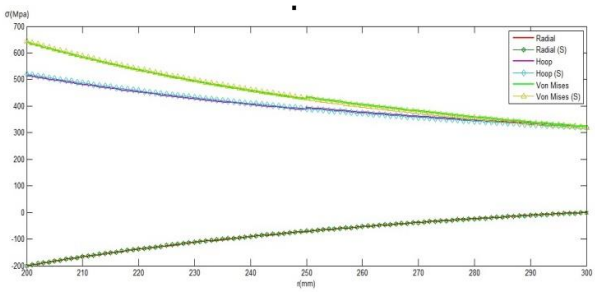
e=0.001



e=0.002

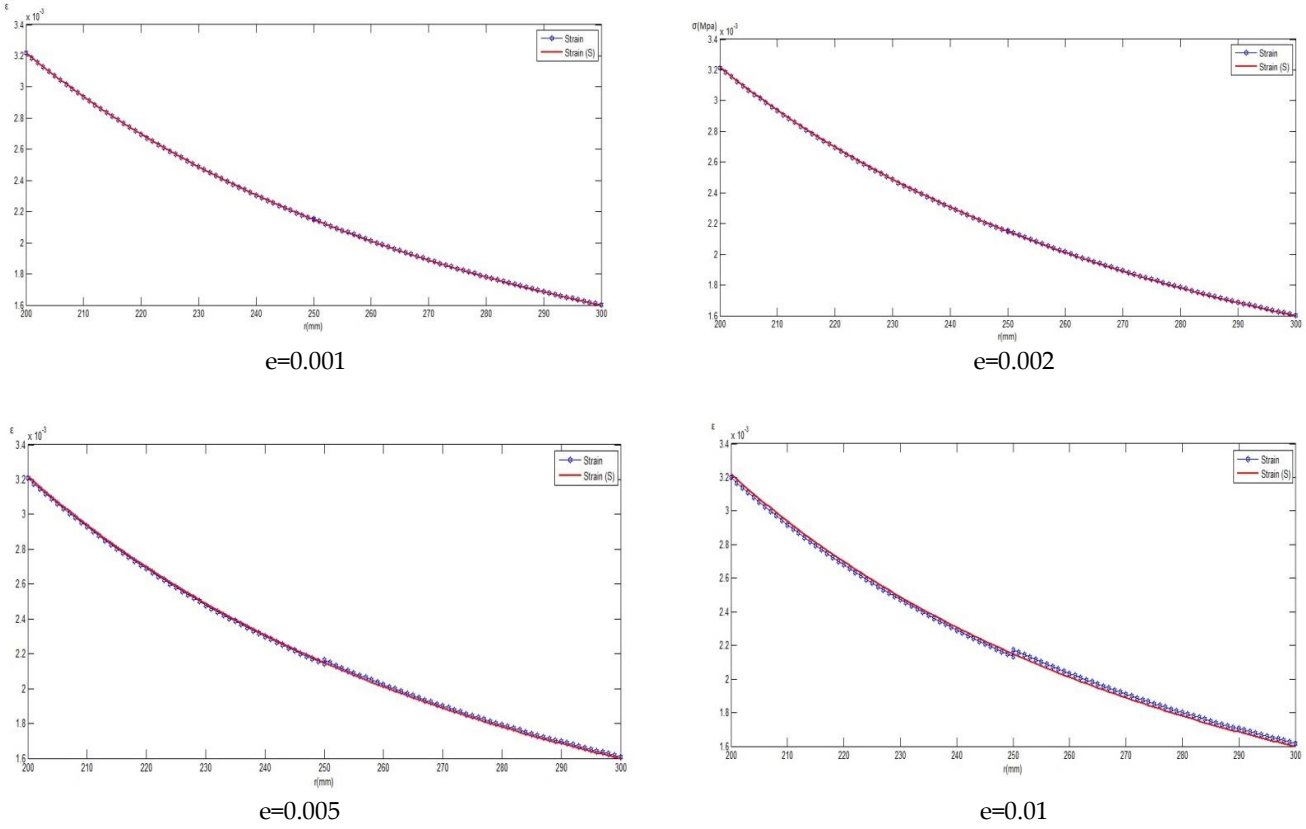


e=0.005



e=0.01

Figure 2. The curve data stresses equations (original and simplified equations(s))



e=0.001 e=0.002

e=0.005 e=0.01

Figure 3. The curve data strain equations (original and simplified equations (S))

III. VALIDATION BY THE FINITE ELEMENT METHOD

The Finite Element Method with ABAQUS Software was used to validate the results of the analytical method using the identical geometrical and mechanical features of the component as in section 2. As in the analytical method, the inner surface of the inner cylinder is exposed to a pressure of 200 MPa without the use of a boundary condition. Fig. 4 and Fig. 5 shows simulation results in the case of e=0.001 for radial, hoop, Von-Mises stress and strain.

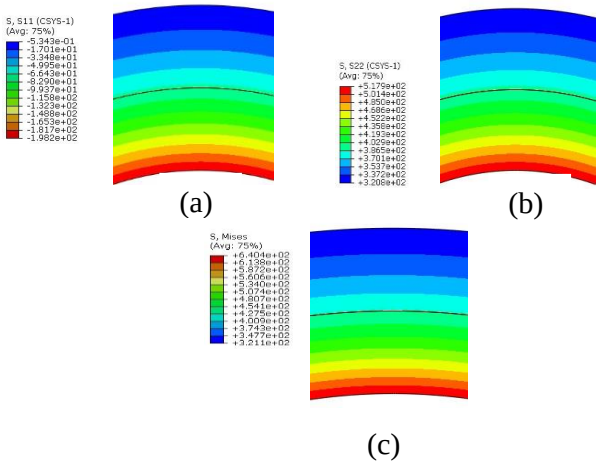


Figure 4. Using the Finite Element Method, radial, hoop, and Von Mises stresses.

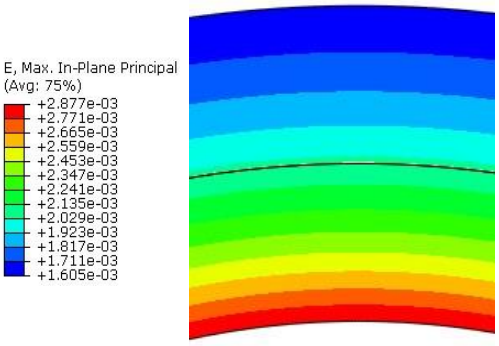
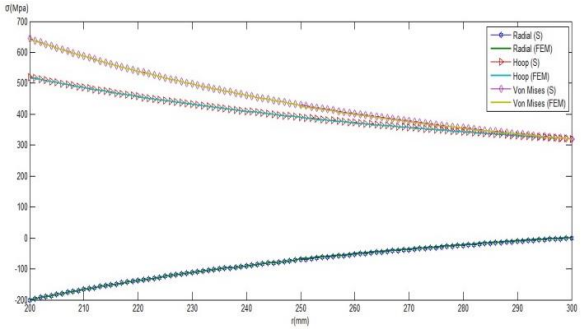


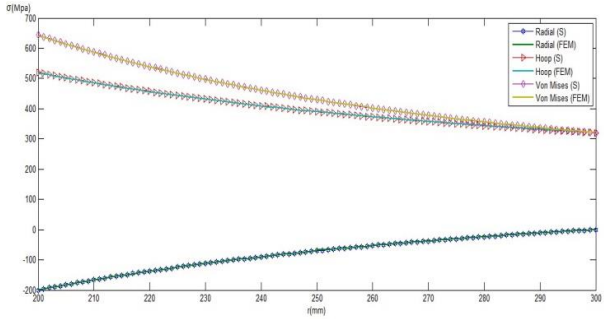
Figure 5 Strain by Finite Element Method

IV. COMPARE THE RESULTS OF SIMPLIFIED EQUATIONS AND FINITE ELEMENT METHOD

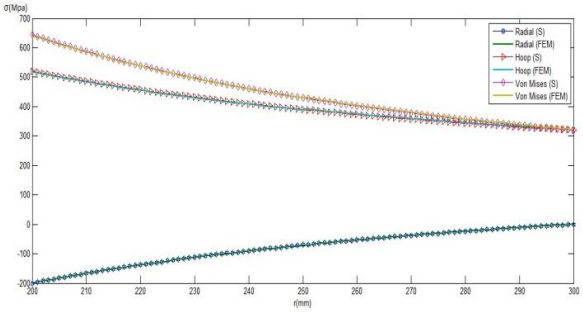
Fig.6. Show how the radial, hoop, and Von Mises stresses are distributed throughout the assembly as a function of the radius r using the simplified equations and the finite element method. It is clear that there is a difference but it is little even at the point of contact between the inner and outer cylinders i.e. at r=250 mm and at the inner and outer surfaces i.e. at r=200 mm and r=300 mm.



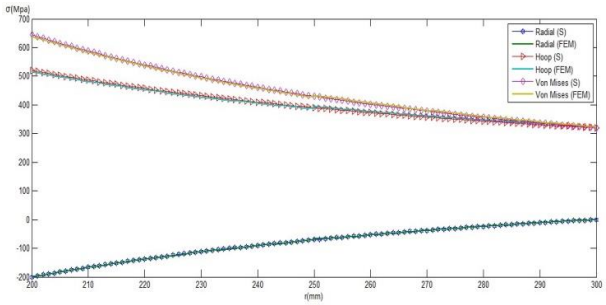
e=0.001



e=0.002



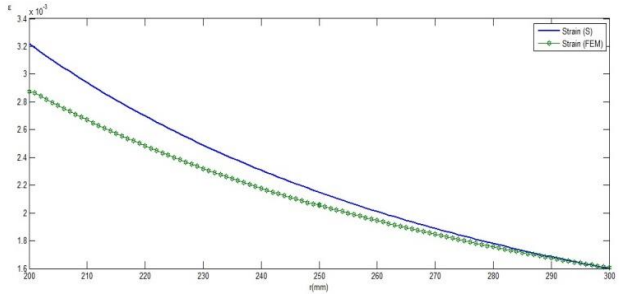
e=0.005



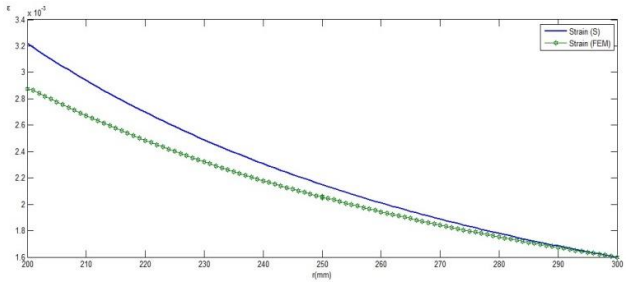
e=0.01

Figure 6. Distribution of radial, hoop and Von Mises stresses a function of radius using FEM

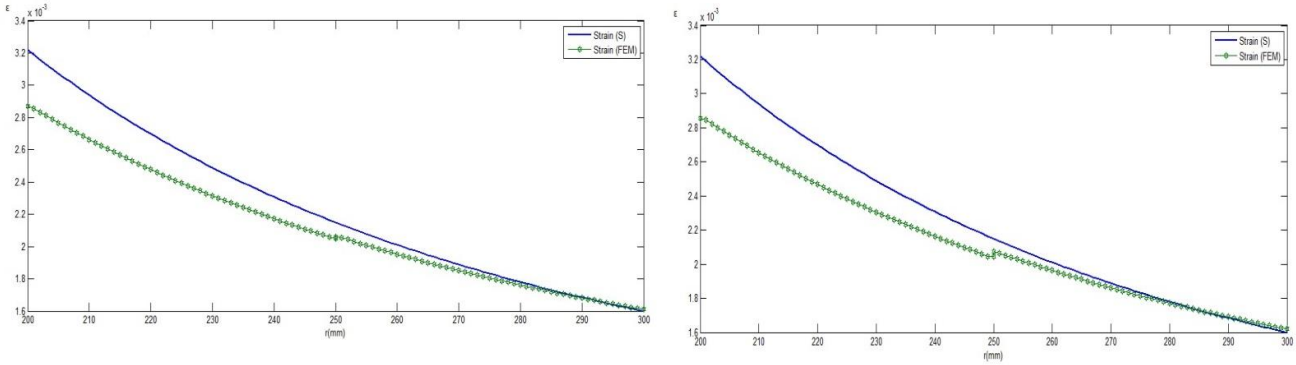
As for the strain, there is a slightly greater difference than the difference in stresses, as shown in Fig. 7



e=0.001



e=0.002



e=0.005 e=0.01

Figure 7. Distribution of strain a function of radius using FEM

V. CONCLUSION

The analytical method was applied using a simplified equation of radial, hoop and Von Mises stresses and strain for a shrink-fit assembly where p is the pressure applied to the inner cylinder's inner surface. This makes the computation of stresses and deformations much easier than it was before. Simulation was used to validate the results as it was demonstrated that the interference value could be mathematically neglected in those cases.

Nomenclature

a	The inner ring's inner radius	r	Radial position
b	The inner ring's nominal outer radius or the outer ring's nominal inner radius	σ_r	Radial stress
c	The outer ring's outer radius	σ_θ	Hoop stress
bi	The inner ring's outer radius	σ_{VM}	Von Mises stress
bo	The outer ring's inner radius	ε	Strain
e	Interference	σ_{rs}	Simplified radial stress
E	Young's modulus	$\sigma_{\theta s}$	Simplified hoop stress
ν	Poisson coefficient	σ_{VMs}	Simplified von Mises stress
p	Pressure	ε_s	Simplified strain

REFERENCES

[1] Alkebsi E A A, Ameddah H, Outtas T and Almutawakel A 2021 Design of graded lattice structures in turbine blades using topology optimization *Int. J. Comput. Integr. Manuf.* 34 370–84

[2] Zhang Y, Mcclain B and Fang X D 2000 Design of interference fits via finite element method 42 1835–50

[3] Laghzale N and Bouzid A 2016 Analytical modelling of elastic-plastic interference fit joints *Int. Rev. Model. Simulations* 9 191–9

[4] Mouâa A, Laghzale N E and Bouzid A-H 2019 Elastic-Plastic Stresses in Shrink Fit with A Solid Shaft *MATEC Web Conf.* 286 02001

[5] Belhaou M and Laghzale N 2019 Analytical Modelling of Elastic Interference Fit Joints for Functionally

Graded Materials *MATEC Web Conf.* 286 02006

[6] Campos U A and Hall D E 2019 Simplified Lamé's equations to determine contact pressure and hoop stress in thin-walled press-fits *Thin-Walled Struct.* 138 199–207

[7] Pedersen P 2006 On shrink fit analysis and design *Comput. Mech.* 37 121–30

Non-coherent fault tree for analysing major risk thermal runaway in adiabatic catalytic fixed-beds reactor

Benomar Fatima
Laboratory of Engineering in Industrial Safety and Sustainable Development "LISIDD" Institut of maintenance and industrial safety.IMSI University of Oran2 Mohamed Ben Ahmed - Algeria.
fbenomar@yahoo.fr

Lounis Zoubida
Laboratory of Engineering in Industrial Safety and Sustainable Development "LISIDD" Institut of maintenance and industrial safety.IMSI University of Oran2 Mohamed Ben Ahmed - Algeria.
lounis_amira@yahoo.fr

Aissani Nassima
Laboratoire d'Ingénierie en Sécurité Industrielle et Développement Durable "LISIDD" Département de Sécurité industrielle & Environnement. IMSI –Université d'Oran 2 Mohamed Ben Ahmed – Algérie
Aissani.nassima@yahoo.com

***Abstract*—**
Thermal runaway is one of the major risks in the chemical industry, occurring when the reaction-generated heat exceeds the heat removed by the cooling system. Such incidents happen frequently, causing copious loss of human life and injury to both present and future generations by malformations. Additionally, thermal runaway causes considerable material and environmental damage. Therefore, this work assesses the cooling system for an adiabatic catalytic fixed-beds reactor using a non-coherent fault tree and importance measures to reveal the weakest areas in the system and improve their reliability and availability. A non-coherent fault tree (nc-FT) is an extension of a coherent fault tree (c-FT) when gate NOT is used as well as gates AND and OR. Where failure state and available stat contribute to system unavailability. This kind of tool (nc-FT) is used when regulation loops are analysed, as occurs in this study's system.

***Keywords*—**
thermal runaway, cooling system, non-coherent fault tree, NOT gate, reliability, availability.

I. Introduction

Chemical and petrochemical industries use reactors for transforming raw materials into consumer goods. A reactor, the core of a plant, is an enclosed vessel where chemical reactions occur. Most of these reactions are exothermic, and it is essential to control this high temperature by implementing a reliable cooling system in the reactor. A cooling system must control the reactor temperature and keep the core of the plant in safety conditions of temperature. If the reaction course is not controlled, the chemical reactor can produce hazardous conditions, like thermal runaway [1]. Uncontrolled temperature causes Thermal runaway which is one of the major risks of the chemical industry, beyond this, thermal runaway causes considerable material and environmental damage [2]. This issue produces 26.5% of petrochemical accidents and 25% of accidents in France [3]. In recent years, this risk has caused many lethal and non-lethal accidents. One such incident was the Seveso disaster, when a toxic cloud was released into the atmosphere through a ruptured disk, poisoning almost 37,000 people [4]. According to Alex Kummer and Tamas Varga, approximately 120 scientific journal articles have been published every year in the last decade on thermal runaway. Most of these studies have examined the thermal properties and process hazard of materials through thermal analysis techniques, such as differential scanning calorimetry (DSC) and an advanced reactive system screening tool (ARSST) [4-6]. Other studies have used thermodynamic calculations and kinetic analysis to assess thermal risk [7, 8]. Additionally, other researchers have used reliability methods like Hazop, FMEA, artificial neural networks and dynamic simulators [9- 13]. The reliability methods are used to study, characterise, measure and analyse the failure and repair of a system to improve their operational use by increasing their design life, eliminating or reducing the likelihood of failure and safety risks and reducing downtime, thus increasing available operating time. Several tools and computers codes must be combined to estimate accident probabilities. Event tree analysis (ETA), fault tree analysis (FTA), reliability bloc diagrams (RBD), Markov models and Bayesian networks (BN) are examples of evaluation tools that can analyse the system safety [14,15]. Beyond identifying weaknesses and critical components in a system, scientists use importance factors to rank components according to their contributions into the probability of failure to the system [16]. This work

assesses the hazardous condition 'loss of cooling system' for a fixed-bed adiabatic catalytic reactor using the fault tree analysis and importance factors. The fault tree (FTA) is a risk analysis method widely used in reliability and safety engineering to examine an undesirable state of a system and identify the best decision to reduce the probability of damage occurring. This strategy is used in various industries, such as nuclear engineering, aeronautics, chemistry, chemical and petrochemicals process, sociology and economics.

Furthermore, this research team uses fault tree analysis in various process, such as the following examples:

- Reliability analysis by mapping probabilistic importance factors into Bayesian belief networks for making decisions in water deluge systems [17],
- Dynamic control for safety system multi agent-system with case- based reasoning [18] and
- Safety assessment of flare systems by fault tree analysis [19].

This paper is organised as follows. Section 2 describes the methodology of a specific non-coherent fault tree. Section 3 explains how this approach is used and produces results. Lastly, Section 4 provides the conclusion.

II. Fault Tree Analysis

Fault tree is a method widely used in fields such as nuclear engineering, aeronautics, finance and chemistry to quantify risks and weak points in a system. The main objectives of the analysis are to determine the possible causes of a specific system failure mode and make decisions [20]. Fault tree analysis (FTA) is a systematic method for acquiring information about a system, the first method developed to conduct a systematic examination of that system's risks. This procedure was introduced in 1962 at Bell Telephone Laboratories after a safety evaluation of the control launching system for intercontinental minuteman ballistic missiles [21]. Beyond this, FTA is a deductive or backward analysis technique that uses a 'What can cause this?' This method is a graphic model composed of interconnected using logic gates. The Fault tree is built from the top representing the system failure named 'Top Event' and developed in branches below this event, show its causes. Each branch of the tree is composed of both intermediate events and basic events.

- Intermediate events are subsystems of the main system, and their failure makes the entire system fail. These events are composed of other subsystems or basic events, depending on the system's complexity.
- Basic events are the origins of accidents or risks and are also called 'root events'. These can include mechanical failures or operator errors.
- Top events, intermediate events and basic events are linked with logical gates like the 'OR' and 'AND' gates.

A. Non-coherent fault tree

A coherent fault tree (c-FT) is described by the fundamental operators (OR and AND) and the monoform variables model the failure of the elements of the system (elementary events of the same category x_i). The complexity of this study's system led the researchers to use other kinds of fault trees, called non-coherent fault trees (nc-FTs) to analyse non-coherent systems. Even improving components reliability (replacement for example), the system remains faulty, and this main characteristic of non-coherent systems is verified in the study system. The latter is a negative feedback control loop. A non-coherent FT contains fundamental operators (OR, AND and NOT; Figure 5) and contains biform variables representing the failure and no-failure of elements of the system (elementary events of different category x_i and $\overline{x_i}$) [22, 23]. In nc-FT, the working component can contribute to system failure. That is, the structure function $\Phi(x)$ of a non-coherent fault tree does not conform to the requirement of coherency.

1) Definition of coherency: A fault tree is coherent if its structure function $\Phi(x)$ complies with the definition of coherency given by two properties [24].

a- Relevance

This first property means that each component is relevant, so each component contributes to the system state.

For example, a system S is composed by n elements or components $(x_1, x_2, \ldots, x_n)$. The status of each component can be shown by an indicator x_i Where $i = 1, 2, 3 \ldots, n$. The structure function $\Phi(x)$ of a system S

$$\Phi = \begin{cases} 1 \text{ if component } i \text{ has failed} \\ 0 \text{ if component } i\text{is working} \end{cases} \quad (1)$$

$x = (x_1, x_2, \ldots, x_n)$

So each component is relevant to the system:

$$\Phi(1_i, x) \neq \Phi(0_i, x) \text{ for some } x_i \quad (2)$$

b– Monotonocity

The second condition for coherency, monotonocity, implies that a structure function, $\Phi(x)$, of the fault tree is monotonically increasing (i.e. non-decreasing).

$$\Phi(1_i, x) \geq \Phi(0_i, x) \quad \forall i \quad (3)$$

Where

$$\Phi(1_i, x) = \Phi(x_1, x_2, ..., x_{i-1}, 1, x_{i+1}, ..., x_n)$$

$$\Phi(0_i, x) = \Phi(x_1, x_2, ..., x_{i-1}, 0, x_{i+1}, ..., x_n)$$

An increasing structure function means that as the component deteriorates or fails the system state either remains the same or also deteriorates. This function produces tree possibilities like those shown in Fig. 1 [24-26].

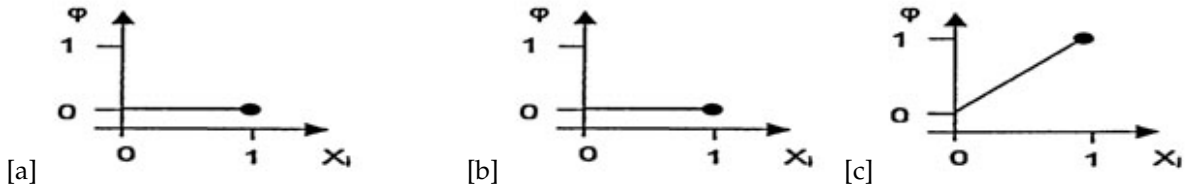


Figure 1. Non-decreasing structure functions

Explication

Fig. 1(a): When $\Phi = 1$, the system is in a failed state due to its other components, while component i is working ($x_i = 0$). The failure of component i ($x_i = 1$) makes no difference to the system state.

Fig. 1 (b): In this case, $\Phi = 0$, the system is functioning and the component i available ($x_i = 0$), and when i fails the system continues to function. This shows that the system is not in critical state for component i.

Fig. 1 (c): This figure shows that the component i is critical for the system; that is, when i fails, the system fails.

It is different in the case of non-coherent system: the system is non-coherent for component i when this component is working the system is in his failed state and then when the component is failed the system is restored to the functioning condition (Fig. 2).

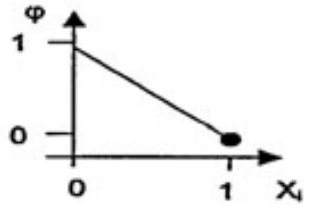


Figure 2. Non-coherent structure function

B. Analysis of non-coherent fault tree

The analysis of a fault tree diagram produces two types of results: qualitative and quantitative.

Qualitative analysis identifies the combinations of basic events (implicants) set in a non-coherent fault tree (nc-FT), which cause the top event. Quantitative analysis is used to calculate the top event probability of the system and to predict its availability and reliability.

1) Qualitative analysis: Qualitative analysis, first, determine the prime implicants set that lead to the main event and then prioritizing these implicants to determine which combination of events could cause the undesirable event (top event).

a-Prime implicants set

Implicants set is a combination of failed components and working components where if they all occur, the top event also occurs. A prime implicants set is a list of events that are both sufficient and necessary for causing the top event. A prime implicants set is the smallest list of component failures and working states where if they all occur, the top event also occurs.

Example for determining prime implicants

Using Fig. 5, one can calculate the prime implicants set for the top event, T.

The top event for second subsystem for the failure of the cooling system $G3$.

$G3 = G3.1 * G3.2$
$G3.1 = EV31 * G31$
$G31 = -EV32$
$G3.1 = EV31 * -EV32$
$G3.2 = EV33 * -EV34$
Then $G3 = (EV31 * -EV32).(EV33 * -EV34)$
The prime implicants for this subsystem are $\{EV31, -EV32\}$ and $\{EV33, -EV34\}$.

2) Quantitative analysis: Quantitative analysis or probabilistic evaluation is a part of the FTA assessment where the probability of the top event must be calculated based on the probabilities of basic events. This process uses Boolean logic which represents the logical gates (OR, AND and NOT).

Additionally, quantitative analysis makes it possible to determine the importance of each component or prime implicants set, making it possible to predict the time to failure. This method also allows one to evaluate the sensitivity of the system to changes in maintenance and inspection time [25].

For the fixed-bed catalytic adiabatic reactor cooling system, all the components are repairable, and inspection tests are conducted every year. Thus, the choice of failure model is the dormant model according to the Isograph software. Isograph software assumes that the dormant components exhibits characteristics similar to non-repairable component during periods between inspections. The actual variation of unavailability is periodic in nature.

Dormant model type

$$Q_{mean} = \frac{\lambda\tau - (1 - e^{(-\lambda\tau)}) + \lambda MTTR(1 - e^{(-\lambda\tau)})}{\lambda\tau + \lambda MTTR(1 - e^{(-\lambda\tau)})} \quad (4)$$

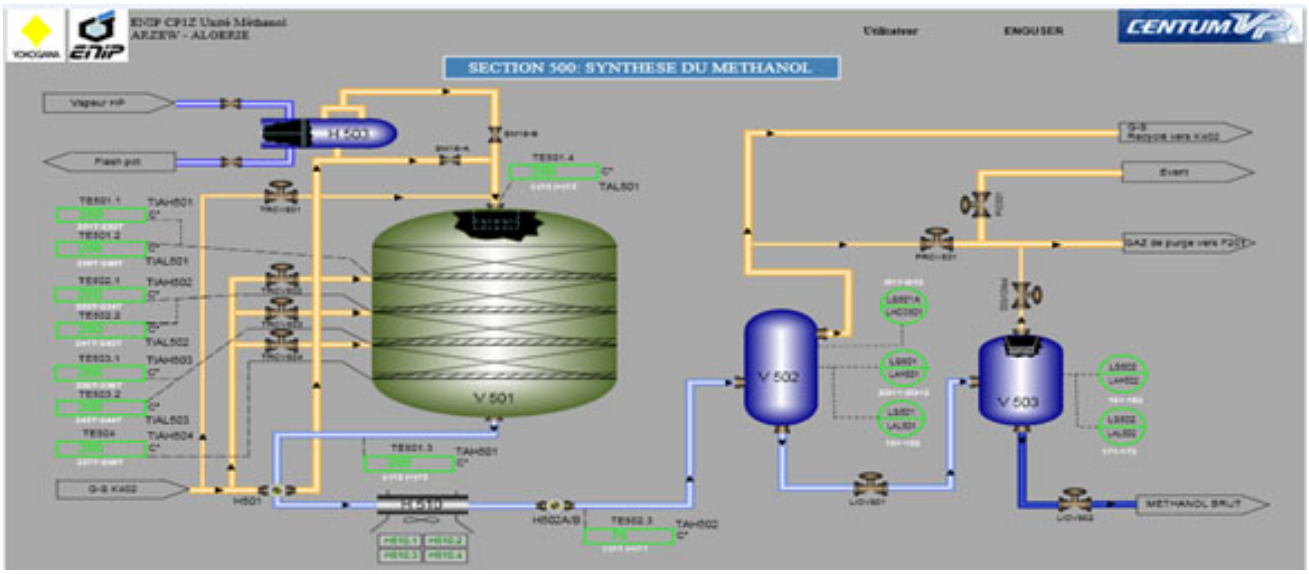


Figure 3. Synthesis loop [27]

For our model we use Qmax for calculating the maximum unavailability.

$$Q_{max} = 1 - e^{-\lambda\tau} \tag{5}$$

Where:
Q_{mean} = Mean unavailability
λ = Constant failure rate
$MTTR$ = Mean time to repair
τ = Test interval (time interval between two consecutive tests).
Q_{max} = maximum unavailability value over the inspection interval.

a-Importance measure

The sensitivity analysis of the system aims to identify its weak points. This analysis makes it possible to find the component or primes implicants set (the minimal cuts set), contributing to the occurrence of undesirable events. This contribution is defined as the importance of component or minimal cuts set. Each component either causes or contributes to the occurrence of the top event and guides the choice of the corrective actions that are most effective in improving its dependability.

Generally, a numerical value is assigned to each basic event, allowing them to be ranked according to the extent of their contribution to the occurrence of the top event [16, 17, and 26].

In this paper, two importance factors are used: Birnbaum and Fussell-Vesely.

III. Case Study

The method of non-coherent fault tree analysis is applied, in this paper, for an adiabatic catalytic fixed-beds reactor for methanol synthesis.

The synthesis loop is the section of the unit where part of the synthesised gas is converted to methanol. In this loop, V501 is a vertical cylindrical container containing four copper-based catalyst beds .

The synthesised gas which enters the reactor V501 goes through the catalyst, here the temperature increases further due to the reaction. The temperature inside the reactor is maintained at a maximum of 270°C by injecting quench gas via the valves (TRC501, TRC502, TRC503 and TRC504).

The synthesised gas is the mixture of recycled gas from the splitter V502 and fresh gas arriving from the flow valve FR402 and enters the synthesis loop via the K402 compressor, where the entered gas is measured using the flow meter FI403 and distributed by FR403.

For temperature control, beyond the TRCs valves, there are four high-temperature indicator alarms (TIAH501, 502, 503, 504) located at the outlets of the four beds and four high temperature sensors (TI541, 542, 543, 544).

A reliable cooling system keeps the reactor under safe conditions, so this work analysed the cooling system for the reactor V501 for the company CP1Z of the SONATRACH group and help to make decisions to improve the RAMs (reliability, availability and maintainability) for this system.

Tab. I shows the different components for the cooling system of the reactor V501 and those components' failures, repair rates, and inspection intervals.

All these values are taken from the database Oreda (Offshore Reliability Data Handbook) [28] except the alarm value, which is taken from the component reliability data for use in the probabilistic safety assessment [29].

Table I. components, their failure and repair rates, and test intervals.

Component		*Failure rate per hour*	*MTTR (hour)*	*Inspection interval(hour)*
FR 402 (Flow valve)	FR402 is stuck closed (EV51)	3.97 e-6	5.3	8760
	FR402 is stuck open(EV52)	7.93 e-6	5.2	
V502 (Splitter) (EV61)		1.07 e-6	2.4	8760
The compressor K402	Mechanical failure (EV71)	9.32 e-6	77.8	8760
	External leak (EV73)	5 e-6	0.5	
FR403 (Flow Valve)	FR403 is stuck closed (EV31)	3.97 e-6	5.3	8760
	FR403 is stuck open (EV32)	7.93 e-6		
FI403 (Flow sensor)	FI403 sensor indicates a high level (EV33)	0.75 e-6	6	8760
	FI403 sensor indicates a low level (EV34)	1.49 e-6	7	
TRC501 (regulating valve for bed1) (EV411)		5.44 e-6	4	-
ST541 (temperature sensor for bed1) (EV415)		2.03 e-6	2	-
TIAH1 (indicator alarm for bed1) (EV413)		7.8 e-7	2	-
TRC502 (regulating valve for bed2) (EV421)		5.44 e-6	4	-
ST542 (temperature sensor for bed2) (EV425)		2.03 e-6	2	-
TIAH2 (indicator alarm for bed2) (EV423)		7.8 e-7	2	-
TRC503 (regulating valve for bed3) (EV431)		5.44 e-6	4	-
ST543 (temperature sensor for bed3) (EV435)		2.03 e-6	2	-
TIAH3 (indicator alarm for bed3) (EV431)		7.8 e-7	2	-
TRC504 (regulating valve for bed 4) (EV439)		5.44 e-6	4	-
ST544 (temperature sensor for bed4) (EV443)		2.03 e-6	2	-
4TIAH4 (indicator alarm for bed4) (EV441)		7.8 e-7	2	-
The turbine has mechanical failure (EV75)		9.32 e-6	77.8	8760
The turbine has an external leak (EV79)		8.47 e-6	48.6	8760
Boiler failure (EV77)		11.98 e-6	16	8760

IV. Result

This study modelled the failure of the cooling system of the adiabatic catalytic fixed-beds reactor with non-coherent fault tree by isograph software [30]. This action produced the results shown in Fig. 4–9. The top event of the tree is "Cooling system failure" chosen the study our system and to analyse how the system is available to keep the balance with the reaction-generated heat and heat removed by the cooling system. According to the structure complexity and to clarify the representation, the FT is split up into six major sub-systems: no synthesis gas flow, compressor system, temperature control failure for bed1, temperature control failure for bed2, temperature control failure for bed3 and temperature control failure for bed4.These systems contribute directly to the top event through an OR gate (G1). The components of the subsystems are listed in table2 and are connected using 'or', 'and' and 'not' gates.

A. Explication of the results

1) Qualitative analysis: Qualitative analysis enables the analyst to determine the prime implicants and qualitative structure importance.

a-Prime implicants set

Based on the previous non-coherent fault trees, this system is composed of 18 prime implicants. Five prime implicants of Order 6, five of Order 4 and 8 of Order 2. Tab. II shows the list for various prime implicants of the cooling system.

The prime implicants are a minimal combination of basic events in their two states, failed and successful, can cause the top event occurrence. These components show where design changes can eliminate or reduce undesirable combinations. Additionally, it is possible to check which specific prime implicants can cause the top event.

In this study's system, there are three subsystems:

1) No synthesis gas (Fig. 4, Fig. 6)
2) No synthesis gas flow (Fig. 5)
3) Temperature control failure (Fig. 7-9)

b-Qualitative structure importance

Qualitative importance gives a qualitative ranking to each subsystem for its contribution to system failure.

For this study's system, the ranking for each subsystem appears in Tab. III.

This ranking is calculated by assuming that each basic event has a probability of 0.001.

$$P(i) = 0.001 \quad \text{then} \quad P(\bar{i}) = 1 - 0.001 = 0.999$$

Tab. III indicates that the weakest subsystem is the temperature control, and the strongest one is the synthesis gas. Similarly, one can qualitatively classify the component for the control temperature subsystem, as demonstrated in Tab. IV. The adiabatic catalytic fixed-bed reactor has four beds, each of which has a sensor, a regulator valve and an alarm to monitor the temperature at that bed.

This qualitative ranking for the subsystem temperature control shows that the regulator valves and alarms are the weakest components in the reactor. The sensors couldn't be ranked at this stage of the study.

2) Quantitative analysis: To quantify the probability of the top event or subsystem of the fault tree, one must have a probability for each basic event (failure rate of

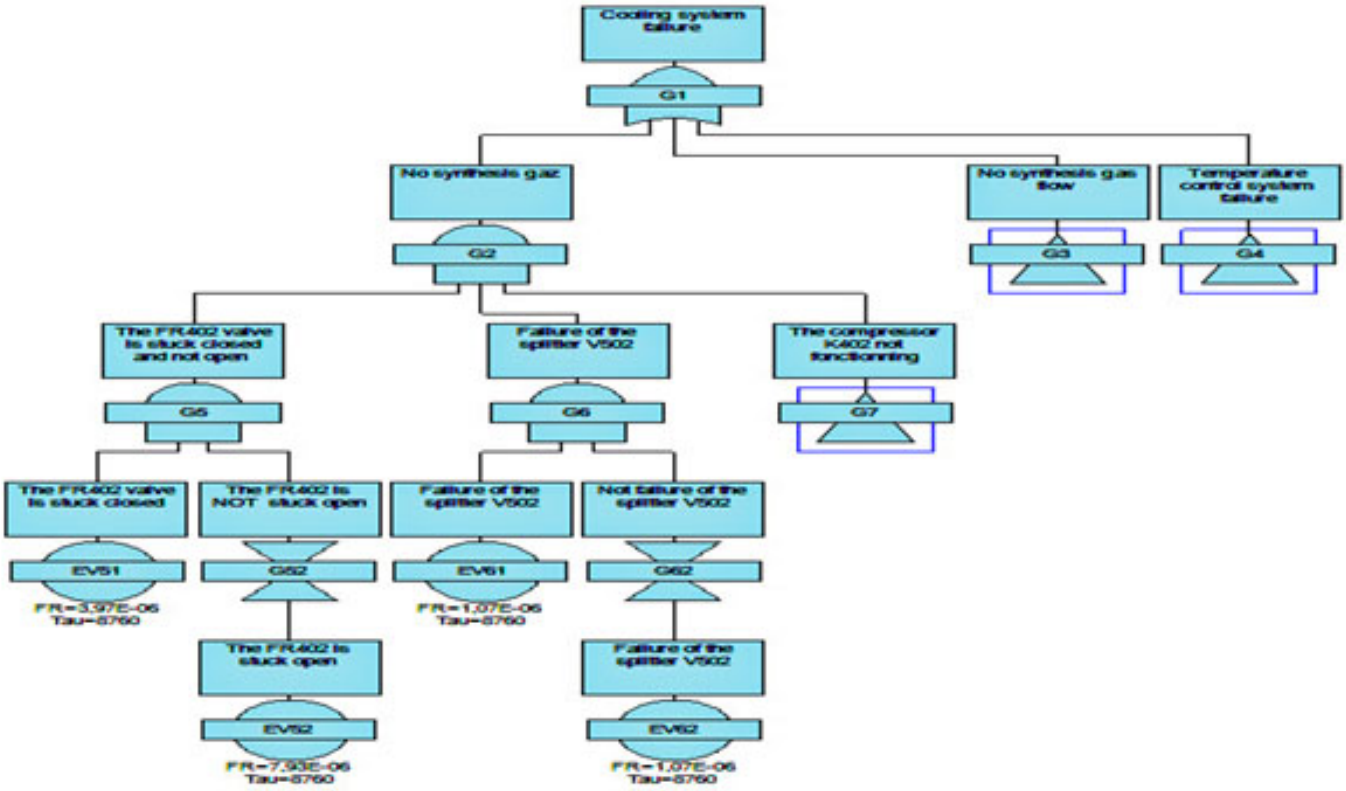


Figure 4. Non coherent fault tree for cooling system failure

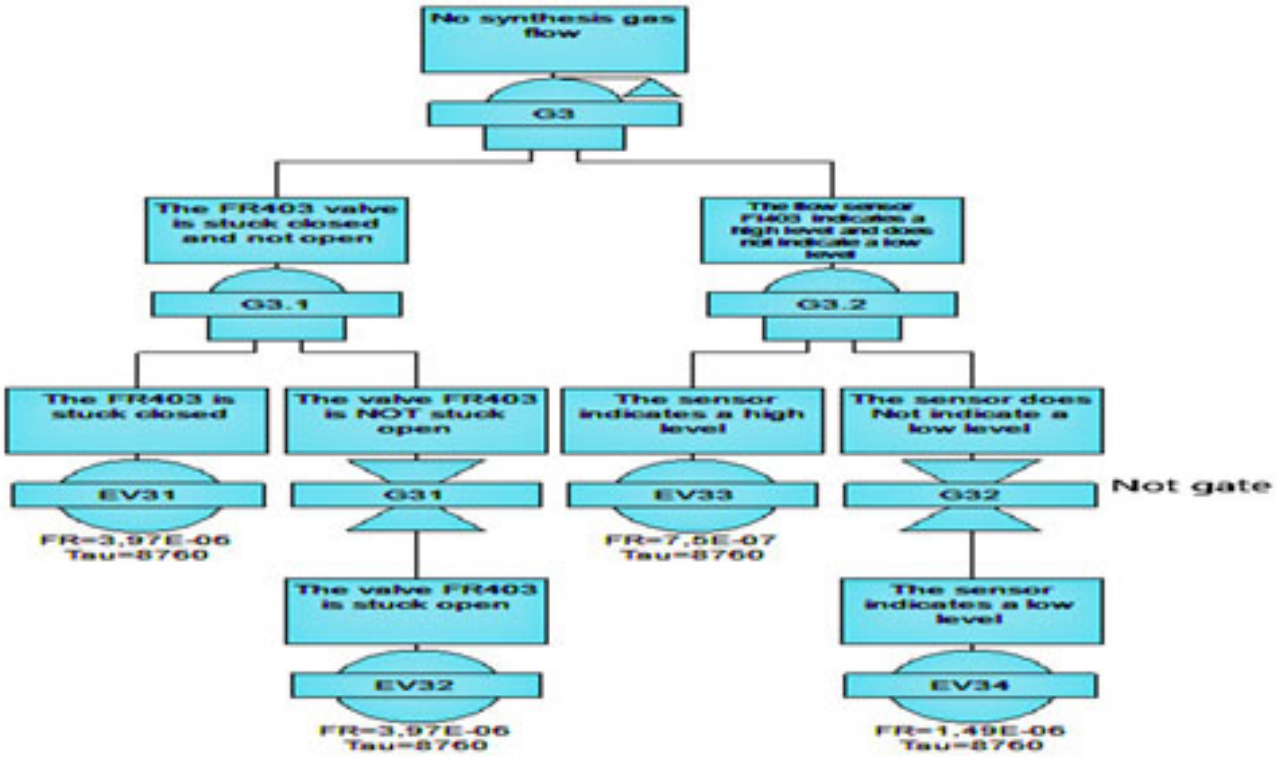


Figure 5. Non-coherent fault tree for subsystem (no synthesised gas flow)

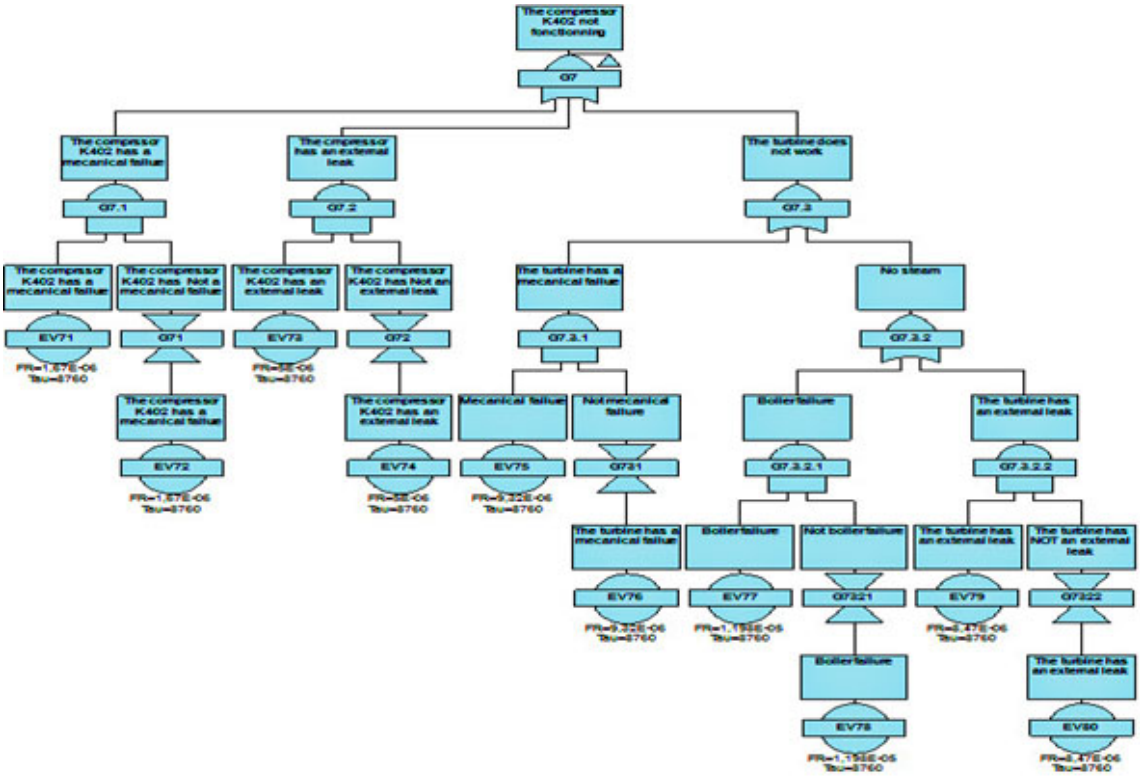


Figure 6. Non-coherent fault tree for compressor

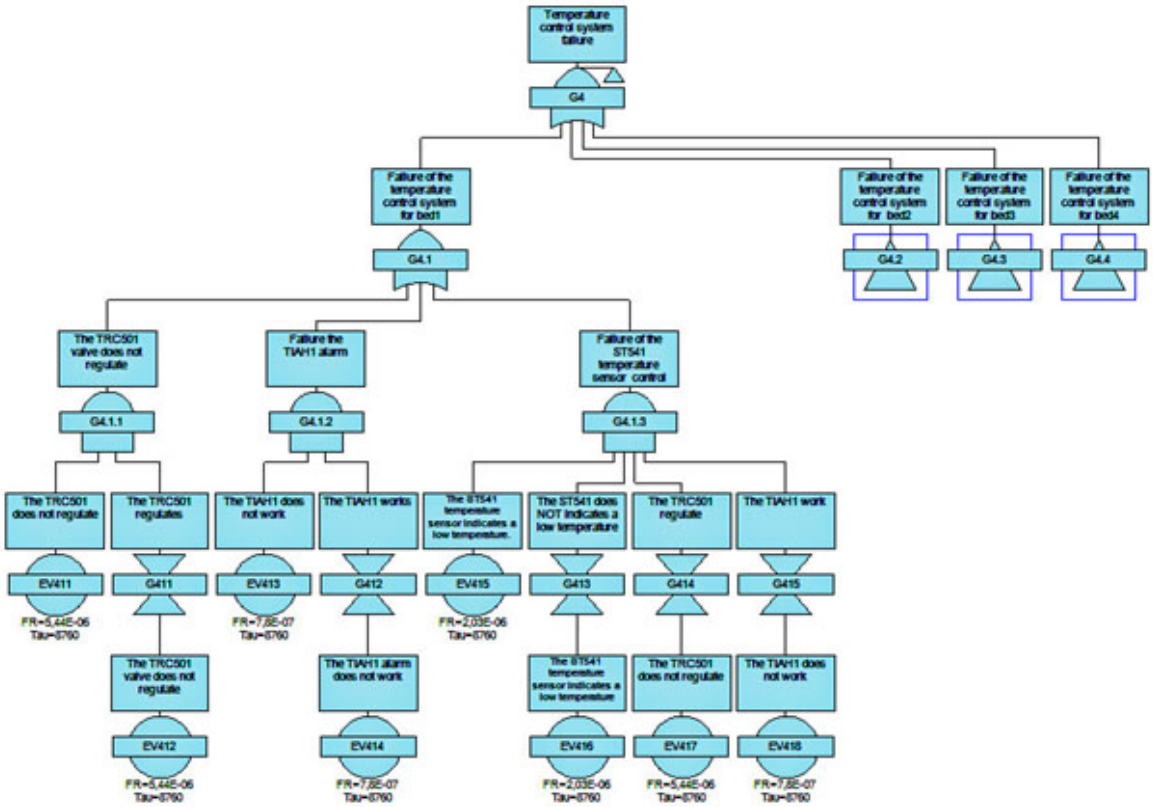


Figure 7. Non-coherent fault tree for subsystem temperature control failure and Bed 1

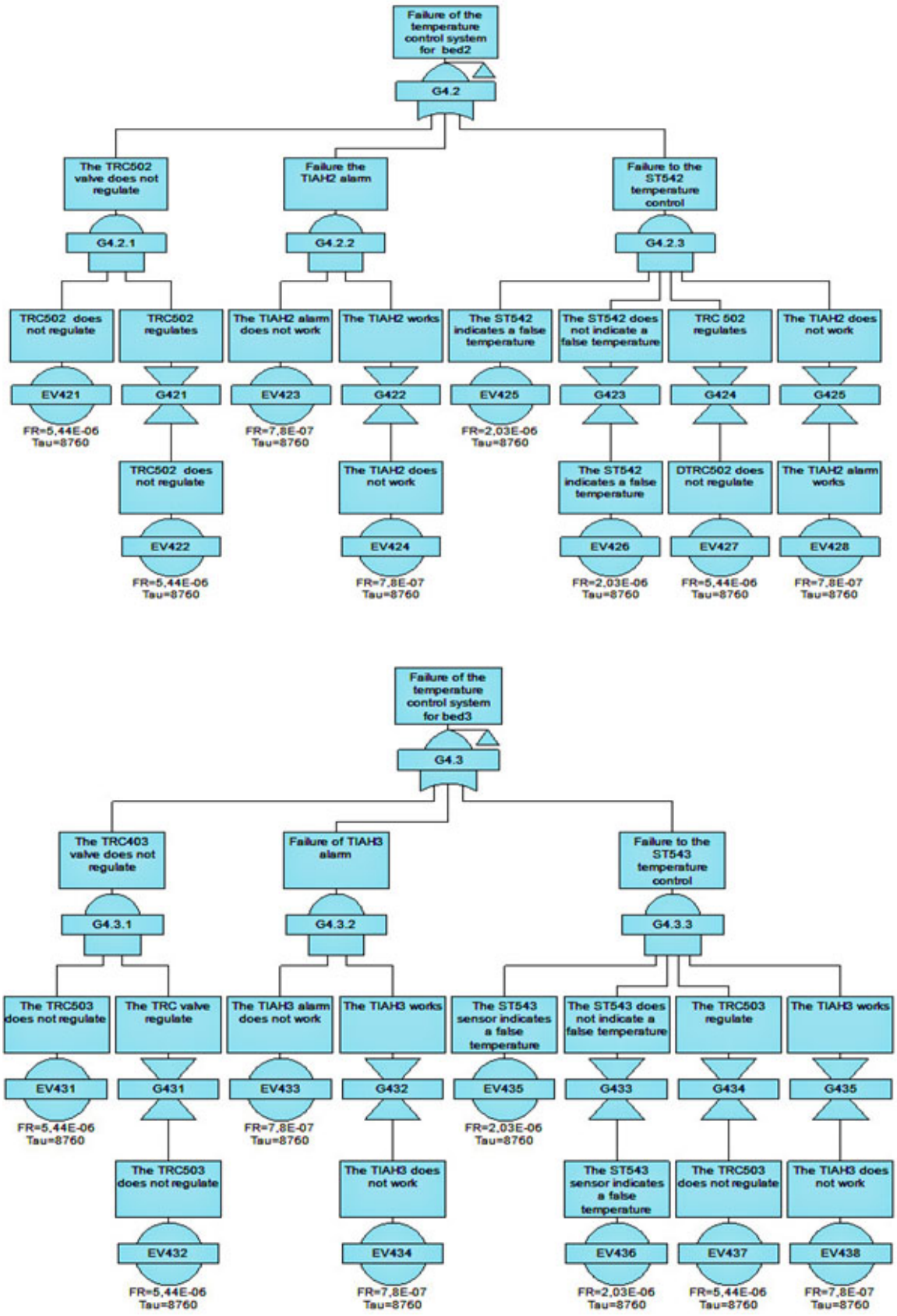


Figure 8. Non-coherent fault tree for subsystem temperature control failure for Beds 2 and 3

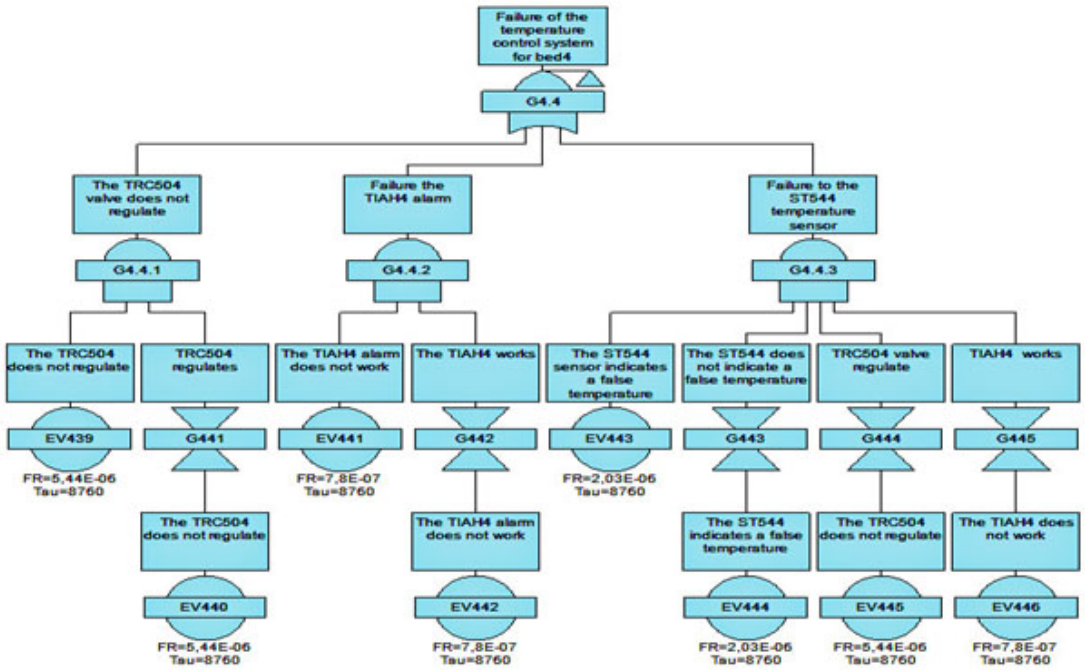


Figure 9. Non-coherent fault tee for failure of temperature control for Bed 4

every component; see Tab. I). Next, these basic component probabilities are propagated upward to the top event using Boolean relationships for the fault tree. This propagation results in defining system parameters like unavailability, reliability, unconditional failure intensity and expected number of failures. Understanding these parameters makes it possible to improve the system and make decisions.

For the system 'Cooling Failure', a quantitative analysis is detailed in Tab. V, and the unavailability for each subsystem is listed in Column 3. Column 5 gives the total number of hours a system will be unavailable in 10 years (87,600 h). The final column ranks the subsystems by their unavailability, and Subsystem 3 (temperature control) has the highest unavailability (0.1312) and 2.5 failures in 10 years.

The unavailability of the Top event (Cooling system failure) calculated by isograph software is as follows:

$$Q = 0.2403$$

Compared with the subsystems' values of unavailability, Subsystem 3's temperature control failure is equal to that of the main event. This subsystem is the most important contributor to system failure.

Additionally, quantitative analysis identifies the components with the greatest effect on system unavailability. Such items can be identified using importance measures that rank the elements quantitatively according to their contributions to system unavailability.

a- Importance measures

Importance analysis is a part of the system quantification process, enabling the analyst to rank each component's contribution to system failure for the cooling system of an adiabatic catalytic fixed-bed reactor. In this study, two importance measures (Birnbaum and Fussell-Vesely) were used to rank each component numerically according to each one's contribution to the cooling system's failure.

b- Birnbaum's measure and its extension for non-coherent analysis

In a coherent system, failure can be caused only by component failure, so in this system, the component can only be failure-critical.

$$I_i^B = \frac{\partial Q_{sys}(t)}{\partial q_i} \tag{6}$$

However, in a non-coherent system, failure is caused by both component failure (i) and component repair $(\bar{i})$. Thus, a component in a non-coherent system can be either failure-critical or repair-critical. On the other hand, i can exist in only one state at a time, so these two states must be considered separately [25].

A component is failure-critical (FC) at time t if and only if the system is in a working state such that the failure of i causes the system to fail.

Table II. Prime implicants for cooling system for adiabatic catalytic fixed-bed reactor

No	*Sub system*	*Prime implicants (PI)*	*Number of 2 PI*	*Number of 4 PI*	*Number of 6 PI*
1	No synthesis gas	(-EV52,EV51,-EV62,EV61,-EV72,EV71) (-EV52,EV51,-EV62,EV61,EV74,EV73) (-EV52,EV51,-EV62,EV61,-EV76,EV75) (-EV52,EV51,-EV62,EV61,-EV78,EV77) (-EV52,EV51,-EV62,EV61,-EV80,EV79)	0	0	5
2	No synthesis gas flow	(-EV32,EV31.-EV34,EV33)	0	1	0
3	Temperature control failure	(-EV412,EV411) (-EV414,EV413) (-EV416,EV417,-EV418,EV415) (-EV422,EV421) (-EV414,EV413) (-EV426,EV427,-EV428,EV425) (-EV432,EV431) (-EV434,EV433) (-EV436,EV437,-EV438,EV435) (-EV440EV439) (-EV442,EV441) (-EV444,EV445,-EV446,EV443)	8	4	0

Table III. Qualitative importance of subsystems

No	*Sub system*	*Number of 2 PI*	*Number of 4 PI*	*Number of 6 PI*	*Probability*	*Rank*
1	No synthesis gas	0	0	5	4.985 e-9	3
2	No synthesis gas flow	0	1	0	9.98e-7	2
3	Temperature control failure	8	4	0	7.992e-3	1

Table IV. Qualitative component importance for temperature control

No	*Prime implicants (PI)*	*PI 2*	*PI 4*	*Probability*
Regulate valve	(-EV412,EV411) (-EV422,EV421) (-EV432,EV431) (-EV442,EV441)	2	0	9.99e-4
Alarm	(-EV414,EV413) (-EV414,EV413) (-EV434,EV433) (-EV442,EV441)	0	0	9.99e-4
sensor	(-EV416,EV417,EV418,EV415) -EV426,EV427,EV428,EV425) (-EV436,EV437,EV438,EV435) (-EV444,EV445,EV446,EV443)	0	4	4.9985e-9

$$I_i^F = \frac{\partial Q_{sys}(t)}{\partial q_i} \tag{7}$$

A component is repair-critical (RC) at time t if and only if the system is in a working state such that the repair of i causes the system to fail. The RC for i is given by $I_i^R(q) = \frac{\partial Q_{sys}(t)}{\partial p_i}$

The analysis of a non-coherent system must consider the roles of FC and RC separately. Subsequently, the overall contribution of i is as follows:

$$I_i^B(q) = I_i^F(q) + I_i^R(q) \tag{8}$$

Where
$I_i^B(q)$: Birnbaum measure of component reliability importance $I_i^F(q)$: component failure-critical $I_i^R(q)$: component repair critical

c- Fussell-Vesely's importance measure and its extension

The extension of this importance factor is the same as that of Birnbaum, considering the FV failure importance and the FV repair importance [25].

The Fussell-Vesely (FV) failure importance = Pr {the failed state of component i contributes to system failure (SF)}.

$$I_{FV_i}^F = \frac{Pr\{U_{K=1,i\in CK}^{np} CK\}}{Q_{sys}(t)} \tag{9}$$

The Fussell-Vesely (FV) repair importance = Pr {the working state of component $i(\bar{i})$ contributes to SF}.

$$I_{FV_i}^R = \frac{Pr\{U_{K=1,\bar{i}\in CK}^{np} CK\}}{Q_{sys}(t)} \tag{10}$$

FV is defined as the prime implicants set containing the negation of the basic event i.

For this study, Tab. VI shows the importance factors of Birnbaum and Fussell-Vesely for an Adiabatic Catalytic Fixed-Beds Reactor.

As indicated by this study's qualitative results, the greatest risk contributors are regulator valves (TRC501, 504); alarms (TIAH1, 4); and temperature sensors (ST541, 544), as shown in Fig. 10-11 and Tab. VII.

Table V. Quantitative system analysis

NO	*Sub system*	*Unavailability*	*Unconditional*	*Expected Down time in 10 years (h)*	*Expected Number of failure in 10 years*	*Rank*
1	No synthesis gas	$1.117\ 10^{(-5)}$	$7.583\ 10^{(-9)}$	0.9781	0.0007	3
2	No synthesis Gas flow	$5.515\ 10^{(-5)}$	$2.537\ 10^{(-8)}$	4.831	0.002	2
3	Temperature control failure	0.3112	$3.298\ 10^{(-5)}$	11490	2.51	1

Table VI. Importance measure

Event	*Fussell-Vesely*	*Birnbaum*	*Event*	*Fussell-Vesely*	*Birnbaum*
EV439	0,1644	0,9765	EV75	2,15E-05	7,42E-05
EV421	0,1644	0,9765	EV79	1,95E-05	7,45E-05
EV411	0,1644	0,9765	EV73	1,17E-05	7,57E-05
EV431	0,1644	0,9765	EV71	5,33E-06	7,66E-05
EV435	0,06116	0,9646	EV72	-5,23E-08	-7,51E-07
EV425	0,06116	0,9646	EV74	-2,58E-05	-1,67E-06
EV443	0,06116	0,9646	EV62	-3,99E-07	-1,19E-05
EV415	0,06116	0,9646	EV80	-7,42E-07	-2,83E-06
EV433	0,02437	0,9966	EV76	-9,05E-07	-3,13E-06
EV423	0,02437	0,9966	EV78	-1,43E-06	-3,93E-06
EV441	0,02437	0,9966	EV34	-2,59E-06	-5,55E-05
EV413	0,02437	0,9966	EV52	-2,99E-06	-5,61E-05
EV33	0,0003954	0,0168	EV32	-6,92E-06	-0,00341
EV31	0,0003954	0,003205	EV414	-8,34E-05	-0,00341
EV51	8,49E-05	0,0006877	EV424	-8,34E-05	-0,00341
EV61	8,49E-05	0,002532	EV434	-8,34E-05	-0,00341
EV77	2,68E-05	7,34E-05	EV442	-8,34E-05	-0,00341
EV446	-0,0002093	-0,00856	EV437	-0,00147	-0,00874
EV418	-0,0002093	-0,00856	EV427	-0,00147	-0,00874
EV438	-0,0002093	-0,00856	EV417	-0,00147	-0,00874
EV428	-0,0002093	-0,00856	EV445	-0,00147	-0,00874
EV436	-0,0005456	0,00861	EV432	-0,003951	-0,02347
EV426	-0,0005456	0,00861	EV422	-0,003951	-0,02347
EV416	-0,0005456	0,00861	EV412	-0,003951	-0,02347
EV444	-0,0005456	0,00861	EV440	-0,003951	-0,02347

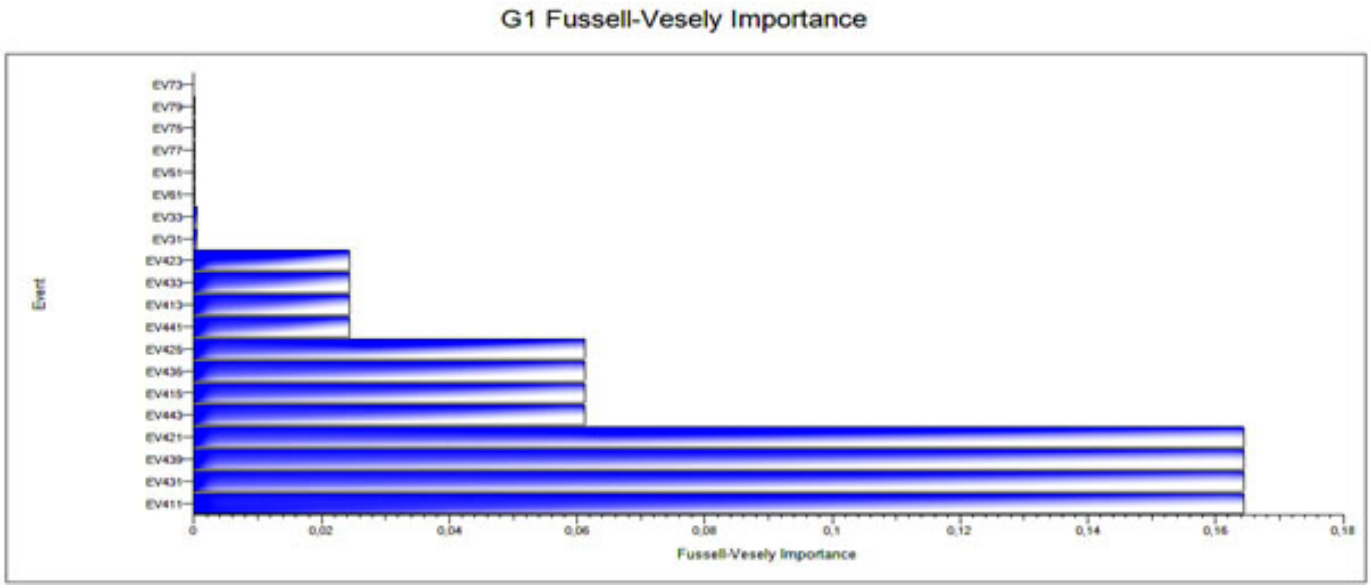


Figure 10. Fussell-Vesely importance measures

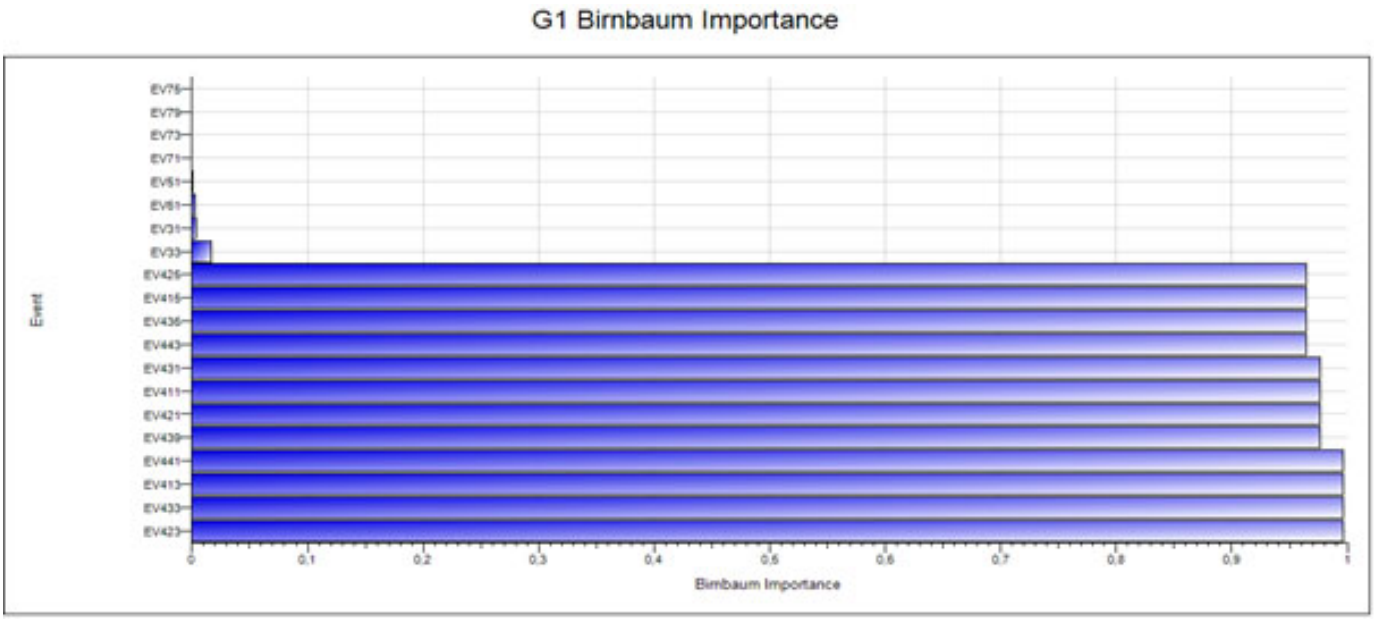


Figure 11. Birnbaum importance measures

Table VII. Ranking of main risk contributors

Event	*Fussell-Vesely*	*Birnbaum*
EV411 (TRC501)	0.1644 (1)	0.9765 (2)
EV421 (TRC502)	0.1644 (1)	0.9765 (2)
EV433 (TRC503)	0.1644 (1)	0.9765 (2)
EV439 (TRC504)	0.1644 (1)	0.9765 (2)
EV415(ST544)	0.06116 (2)	0.9646 (3)
EV425(ST541)	0.06116 (2)	0.9646 (3)
EV435(ST543)	0.06116 (2)	0.9646 (3)
EV441(ST542)	0.06116 (2)	0.9646 (3)
EV413(TIAH1)	0.02437 (3)	0.9966 (1)
EV423(TIAH2)	0.02437 (3)	0.9966 (1)
EV433(TIAH3)	0.02437 (3)	0.9966 (1)
EV443(TIAH4)	0.02437 (3)	0.9966 (1)

V. Conclusion

The fault tree is a practical tool for analysing complex systems and improving their reliability and availability. However, some systems are not coherent, like the one in this study (the cooling system for the adiabatic catalytic fixed-beds reactor). This system used a non-coherent fault tree as an extension for a coherent fault tree, where the working state and the failing state of component were used to analyse the risk of thermal runaway.

In this paper, the results of the non-coherent fault tree for the cooling system predicted a probability of failure equal to 0.1313. In Sub-system 3, the temperature control had the probability 0.1312, almost equal to that of the main event. Sub-system 3 is the biggest contributor to the risk of the loss of cooling and onset a thermal runaway.

The Birnbaum and Fussell-Vesely importance factors showed that the components of Sub-system 3 were the most critical. The Birnbaum factor produced almost same criticality value, whereas the Fussell-Vesely ranked control valves first. Therefore, to improve the reliability and availability for this system, it would be necessary to improve the reliability of the three components of Sub-system 3, the valves design should be changed or the inspection interval time reduced.

Further research could explore the future reliability estimation. Many predictive methods could be tested, such as the Bayesian approach.

References

[1] F. Stoessel, "*Thermal Safety of chemical Processes: Risk Assessment and Process Design*," Wiley-VCH, 2008.

[2] R. Ball, "*Some aspects of theory and mathematics of thermal runway in reacting chemical systems that may be relevant to critical behaviour in confined plasmas*," Proc.R.Soc.Lon., 1999.

[3] A. Dakkoune, L. Vernières-Hassimi, D lefebvre and E. Lionel'*Early detection and diagnosis of thermal runaway reactions using model-based approaches in batch reactors*," Elsevier, 2020.

[4] A. Kummer and V. Tamas, "What do we know already about reactor runaway?- A review," *Process Safety and Environmental Protection*, 2021, 147, pp 460-476.

[5] C. Villemur, L. Petit, N. Bianchini and P. Rotureau, "Runaway Reaction Hazard Assessment for Chemical Process Safety," *Chemical Engineering Transactions*, 2019, 77, pp 451-456.

[6] L. Vernières-Hassimi andS. Leveneur, "Alternative method to prevent thermal runway in case of error on operating conditions continuous reactor," *Process Safety and Environmental Protection*, 2015, 98, pp 365-373.

[7] Y. Kouhili, L. Vernières-Hassimi and L. Estel, "Performance of Runaway Detection in a Batch Reactor using Thermal runaway Criteria," *Chemical Engineering Transactions*, 2022, 91, pp 553-558.

[8] S. Copelli, M. Derudi, C.S. Cattaneo, G. Nano, M. Raboni, V. Torretta and R. Rota, "Synthesis of 4-Chloro-3-nitrobenzotrifluoride: Industrial thermal runaway simulationdue to cooling system failure," *Process Safaty and Environmental Protection*, 2014, 92, pp 659-668.

[9] F. Berdouzi, "*Simulation dynamique de dérives de procédés chimiques. Application à l'analyse des risques [Dynamic simulation of chemical process drifts. Application to risk analysis]*," PhD thesis, University of Toulouse, 2017.

[10] F. Berdouzi, C. Villemur, N. Olivier-Maget and N. Gabas, "Dynamic Simulation for Risk analysis: Application to an exothermic reaction," *Process Safety and Environmental protection*, 2018, 113, pp 149-163.
[11] B. Benamrane, N. Bourmada, and Y. Chetouani, "Analysis of thermal runway scenarios in a chemical reactor," *J.Chemical Engineering transactions*, 2011, 25, pp 255-260.
[12] S. Ardi, H. Minowa and K. Suzuki, "Detection of runway reaction in a polyvinyl chloride batch process using artificial neural networks," *International Journal of Performability Engineering* , 2009, 15(4), pp 367-376.
[13] D. Rizal, S. Tani, K. Nishiyama and K. Suzuki, "Safety and reliability in a polyvinyl chloride batch process using dynamic simulator-case study: Loss of containment incident," *Journal of Hazardous Materials*, 2006, 137(3), pp 1309-1320.
[14] A.L. Raso, V. Vasconcelos, R.O. Marques, W.A. Soares and A.Z. Mesquita, "Use of Reliability Engineering Tools in Safety and Risk Assessment of Nuclear Facilities," *International Nuclear Atlantic Conference - INAC 2017*, Belo Horizonte, MG, Brazil, 2017.
[15] F. Khan, S. Rathnayaka and S. Ahmed, "Methods and models in process safety and risk management: Past, present and future," *Process safety and environmental protection*, 2015, 98, pp 116-147.
[16] M. Modarres, "*Risk Analysis in Engineering: Techniques, Tools, and Trends*," Taylor & Francis Group, 2006.
[17] I.H.M. Guetarni, N. Aissani, E. Châtelet and Z. Lounis, "Reliability analysis by mapping probabilistic importance factors into Bayesian belief networks for making decision in water deluge system," *Process Safety Progress*, 2018, 38(2), e12011.
[18] N. Aissani, I.H.M. Guetarni and S. Zebirate, "Dynamic control for safety system multi-agent system with case based reasoning," *International Journal of Reliability and Safety*, 2017, 11(3-4), pp 238-255.
[19] M.T. Berrouane and Z. Lounis, "Safety assessment of flare systems by fault tree analysis," *Journal of Chemical Technology and Metallurgy*, 2016, 51(2), pp 229-234.
[20] E.E. Hurdle, "*System Fault Diagnosis Using Fault Tree Analysis*," PhD thesis, Loughborough University, 2018.
[21] J.F. Sihite, "*Failure analysis of Power Transformer Based on Fault Tree Analysis*," PhD thesis, Kyoto University, 2013.
[22] R. Ziani, "A Fast Algorithm for Non-coherent Fault Tree Analysis," *Courrier du savoir*, 2007, 8, pp 111-116.
[23] N. Limnios, "*Fault tree*," ISTE Ltd, 2007.
[24] S.C. Beeson, "*Non-Coherent Fault Tree*," PhD thesis, Loughborough University, 2002.
[25] S. Beeson and J.D. Andrews, "Importance Measures for Non-Coherent-System Analysis," *IEEE Transactions on Reliability*, 2003, 52(3), pp 301-310.
[26] J.D. Andrews, "To not or not to not," *in Proceeding of the 18th International System Safety Conference*, 2000.
[27] CP1Z company history report.
[28] C. Oreda, "*Offshore Reliability Data Handbook, 4th ed*," Norway: Det Norske Veritas (DNV), 2002.
[29] Component Reliability Data For use in Probabilistic Data Safety Assessment. A technical document issued by the international atomic energy agency, Vienna, 1988. IEAEA-TECDOC-478.
[30] Isograph, Reliability workbench, Version 14.0.

Etude fiabiliste de la stabilité des talus rocheux

Mekhazni Radia
Laboratoire de recherche en hydraulique Appliquée et Environnement, Université de Bejaia
radia.mekhazni@univ-bejaia.dz

Sadaoui Omar
Laboratoire du génie de la construction et de l'architecture, Université de Bejaia
omar.sadaoui@univ-bejaia.dz

Maza Mustapha
Laboratoire de recherche en hydraulique Appliquée et Environnement, Université de Bejaia
mustapha.maza@univ-bejaia.dz

Résumé—

Actuellement, les méthodes probabilistes et fiabilistes sont largement appliquées dans l'étude de la stabilité des massifs rocheux. Ces méthodes permettent l'estimation des incertitudes et de la fiabilité des modèles déterministes. Le présent article s'intéresse à l'application des méthodes des équilibres limites (méthode de Bishop) et des éléments finis pour le calcul d'un facteur de sécurité unique. Il en est de même pour le calcul d'un facteur de sécurité moyen et d'une probabilité de rupture par l'application des deux approches probabilistes : la simulation de Monte-Carlo et la méthode d'approximation ponctuelle de Rosenblueth. Cet article évoque une étude comparative entre les analyses déterministes et probabilistes de la stabilité de talus rocheux des gorges de Kherrata au PK (point kilométrique) 1+172m.

Mots-Clés—

Talus rocheux, gorges de Kherrata, méthode de Bishop, méthodes des éléments finis, la simulation de Monte-Carlo, méthode d'approximation ponctuelle de Rosenblueth.

I. INTRODUCTION

Les problèmes de la stabilité des talus rocheux sont des phénomènes fréquents qui touchent directement les pentes naturelles et artificielles. Plusieurs cas d'éboulement ont été signalés dans différentes zones rocheuses de la wilaya de Béjaia dont les plus récents sont : l'éboulement de Cap Carbon en 2021 et l'éboulement au niveau de l'ancien tunnel menant à la ville d'Aokas en 2015.

La stabilité des talus rocheux est généralement caractérisée par des méthodes de conception déterministes comme les méthodes des équilibres limites et les méthodes des éléments finis (MEF). Ces méthodes permettent le calcul d'un facteur de sécurité (FS), qui est défini comme le rapport entre les forces résistantes moyennes et les forces motrices moyennes sur une surface de glissement potentielle [1].

Cependant, l'hétérogénéité des pentes rocheuses et les incertitudes liées aux essais in-situ et de laboratoire rendent les résultats des méthodes déterministes peu fiables. Il est donc recommandé d'utiliser des méthodes probabilistes pour évaluer la fiabilité des méthodes déterministes.

L'analyse de la stabilité par les méthodes probabilistes se base sur un système aléatoire, où l'occurrence d'une rupture de talus rocheux est un événement aléatoire dépendant de la valeur aléatoire du facteur de sécurité et des paramètres d'entrées aléatoires géotechniques de la roche en question.

Les méthodes fiabilistes évaluent les incertitudes liées au facteur de sécurité par un indice de fiabilité [2]. Ces incertitudes sont évaluées par différentes approches probabilistes telles que la méthode d'approximation ponctuelle de Rosenblueth et la méthode de simulation de Monte-Carlo.

Cet article présente une étude comparative des résultats de la simulation de Monte-Carlo combinée avec la méthode de Bishop et la méthode d'approximation ponctuelle de Rosenblueth combinée avec la MEF afin de vérifier la stabilité du talus rocheux des gorges de Kherrata au PK 1+172m.

II. METHODOLOGIE

Les analyses probabilistes de la stabilité des talus rocheux sont effectuées par la combinaison des méthodes des équilibres limites, des éléments finis aux méthodes probabilistes. Dans cette étude, la simulation de Monte-Carlo est combinée à la méthode d'équilibre limite de Bishop, Celle d'approximation ponctuelle de Rosenblueth est assemblée avec MEF.

La méthodologique suivie dans cet article se résume en quatre points :

- Définir un modèle géométrique ;
- Déterminer les valeurs uniques et statistiques des paramètres géotechniques de la roche ;
- Effectuer des études déterministes et des études probabilistes ;

- Déterminer la valeur unique et moyenne de facteur de sécurité, l'écart type de facteur de sécurité moyen, la probabilité de rupture et l'indice de fiabilité.

III. LES METHODES DETERMINISTES

A. *Les méthodes des équilibres limites*

L'étude de la stabilité des talus rocheux qui présentent des ruptures (planes, circulaires, en dièdres) dans des régions géologiques complexes, prend un temps considérable dans l'analyse du facteur de sécurité. Pour faire face à la contrainte de temps, une série d'hypothèses simplificatrices sont formulées afin de déterminer une seule valeur de FS.

En raison de différentes hypothèses, diverses techniques ont été développées telles que les méthodes des équilibres limites. Toutes les méthodes dites méthodes des équilibres limites reposent sur des hypothèses simplificatrices basées sur la comparaison des forces résistantes et des forces motrices [2].

La méthode d'équilibre limite la plus utilisée est celle de Bishop simplifié [3]. Elle est appliquée sur une surface de glissement circulaire et permet de déterminer un moment global.

Dans cette étude, l'analyse de la stabilité des pentes par la méthode d'équilibre limite est effectuée au moyen du code informatique Slide 2D V.6.0 (Rocscience Inc 2012). Le programme a également réalisé une analyse probabiliste pour l'évaluation de la stabilité du talus rocheux.

1) *La méthode de Bishop simplifiée*

C'est une technique d'analyse déterministe qui prend en compte des surfaces de glissement circulaire, elle est appliquée dans le cas des roches tendres ou très broyées. Cette technique est fondée sur le principe de la méthode des tranches et elle consiste à découper le talus rocheux en tranches verticales. La méthode de Bishop ne néglige pas les forces horizontales inter-tranches mais la résultante des forces verticales est nulle, d'où : Xi = Xi+1 mais Ei ≠ Ei+1[4]. Cette méthode nécessite un processus itératif et il est programmer par ordinateur. On suppose que la première valeur du facteur de sécurité d'entrée est égale au FS de Fellenius, le coefficient de sécurité calculé est utilisé dans la prochaine itération et ainsi de suite. La formule du facteur de sécurité est donnée comme suit [5] :

$$FS = \frac{1}{W \sin\alpha} \sum \frac{\left[c'b + (W - ub\tan\varphi)\right]\cos^{-1}\alpha}{1 + \frac{\tan\alpha \tan\varphi'}{FS}} \qquad (1)$$

B. *La méthode des éléments finis*

La MEF est une méthode robuste et capable d'évaluer les problèmes complexes de la stabilité des talus rocheux. Elle est basée sur la mécanique des milieux continus, divise le domaine de modélisation généralement en éléments triangulaires liées par des nœuds.

Il existe plusieurs approches pour l'analyse de la stabilité des pentes à l'aide de la MEF, dont l'approche la plus courante est celle de la réduction de la résistance au cisaillement [6], [7].

1) *L'approche de la réduction de la résistance au cisaillement*

Cette approche est largement utilisée dans la modélisation numérique et elle présente de bons résultats comparativement aux méthodes conventionnelles et elle ne requiert pas une évaluation des modes et des mécanismes de rupture [8]. Le FS est calculé par la réduction de la résistance au cisaillement jusqu'à la rupture.

Le calcul de la réduction de la cohésion et l'angle de frottement du talus rocheux se fait par l'application des formules ci-dessous [9] :

$$C_f = \frac{C}{SRF} \qquad (2)$$

$$\phi_f = \tan^{-1} SRF\left(\frac{\tan\phi}{SRF}\right) \qquad (3)$$

SRF (facteur de réduction de la résistance) présente aussi la valeur de FS.

C_f , ϕ_f sont des paramètres plastiques de la résistance au cisaillement.

Dans cet article, l'étude de la stabilité de talus des gorges de Kherrata est effectuée par l'approche de réduction de la résistance au cisaillement qui se repose sur la méthode des éléments finis. L'analyse par la méthode des éléments finis est menée par le logiciel Phase 2D V.8.0 (Rocscience Inc 2015). Le programme a aussi procédé à une analyse probabiliste.

IV. LES METHODES PROBABILISTES

A l'heure actuelle, les ingénieurs en géotechnique et des mines utilisent les méthodes probabilistes dans les études de la stabilité des talus rocheux afin de prédire précisément la variabilité des propriétés mécaniques et physiques des roches et évaluer leurs incertitudes.

L'évaluation de ces incertitudes par les méthodes probabilistes est estimée par l'indice de fiabilité, la probabilité de rupture (*Pf*) et les moments statistiques (la valeur moyenne μ et l'écart type σ) de FS.

Il existe plusieurs méthodes numériques probabilistes permettant d'avoir des résultats fiables de FS pour une bonne étude de stabilité et une meilleure conception de soutènement. On distingue la simulation de Monte-Carlo et la méthode d'approximation ponctuelle de Rosenblueth.

A. La simulation de Monte-Carlo

La méthode de Monte-Carlo est très puissante et flexible, elle peut être appliquée pour résoudre des grands problèmes de stabilité. Cette simulation est très simple à utiliser et elle est précise [10].

Trois étapes sont généralement nécessaires dans la simulation de Monte-Carlo [2] :

- Déterminer les variables indépendantes (variables d'entrées),
- Transformer les variables d'entrées indépendantes en variables de sorties,
- Analyser les variables de sorties.

Dans cette simulation les paramètres statistiques des variables d'entrée sont :

- La moyenne qui est définie par la formule ci-dessous :

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{4}$$

- L'écart type est calculé par cette formule :

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{5}$$

- Le maximum et le minimum de la valeur aléatoire dans une distribution probabiliste normale, environ 99% des échantillons, doivent se situer à l'intérieur de trois écarts types de la moyenne.

Dans cet article le maximum et le minimum de la valeur aléatoire sont égaux à trois écarts types.

Les variables de sortie de la simulation de Monte-Carlo comprennent :

- La probabilité de rupture : Elle est le rapport de nombre d'échantillons d'un facteur de sécurité inférieur à 1.5 sur le nombre total d'échantillons.

$$PF = \frac{\mathrm{N(échantillons)} < 1,5}{N_{total}\,(échantillons)} \tag{6}$$

- Indice de la fiabilité : L'indice de fiabilité est une indication de la valeur d'écart type qui sépare le facteur de sécurité moyen du facteur de sécurité critique. On suppose que le facteur de sécurité est normalement distribué et l'expression de fiabilité normale est donnée par cette formule :

$$\beta = \frac{\mu_{Fs} - 1}{\sigma_{Fs}} \tag{7}$$

β : Indice de fiabilité,

μ_{Fs} : Moyenne de facteur de sécurité,

σ_{Fs} : Ecart type de facteur de sécurité.

L'équation ci-après est utilisée pour calculer l'indice de fiabilité lognormale :

$$\beta_{LN} = \frac{\ln\left[\frac{\mu}{\sqrt{1+V^2}}\right]}{\sqrt{1+V^2}} \tag{8}$$

μ : Moyenne de facteur de sécurité,

V : Coefficient de variation de FS($v = \frac{\sigma}{\mu}$).

B. La méthode d'approximation ponctuelle de Rosenblueth

Cette procédure d'estimation ponctuelle a été proposée à l'origine par Rosenblueth (1975). Les fonctions de densité de probabilité des variables aléatoires sont simulées par des points ponctuels combinés à des pondérations afin d'obtenir une approximation de la distribution de la solution. La méthode d'approximation ponctuelle de Rosenblueth mis en œuvre dans Phase 2D V.8.0 (Rocscience Inc ,2016) est une méthode d'estimation en deux points pour le premier et le second moment des variables non corrélées. Elle nécessite 2^n évaluation présentant le nombre de variables aléatoires.

La moyenne et l'écart type de la solution $y = f(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_n)$ sont donnés par les formules ci-dessous [11]:

$$\overline{y} = \sum_{i=1}^{2^n} wf_i = \sum_{i=1}^{2n} P_{i,\mathrm{j,k,l,n}} \cdot \mathrm{y}_{i,j,k,l,n} \tag{9}$$

$$\sigma_y = \sqrt{\sum_{i=1}^{2^n} wf_i^{\,2} - \left[\sum_{i=1}^{2^n} wf_i\right]^2} \tag{10}$$

La pondération (w) représente le (P) qui est la probabilité correspondante à chaque variable x .

Les suites i, j, k, l, n se sont des permutations de signe $\pm$ pour la combinaison 2^n.

Comme les variables sont symétriques et non corrélées alors w est donnée par $1/2^n$.

f_i est une évaluation successive de f par la combinaison 2^n.

V. LA DESCRIPTION DE LA ZONE D'ETUDE

La zone d'étude se trouve à environ 65 Km au Sud-Est de Bejaia, à 43km au Nord-Ouest de Sétif et à 7 km de la ville de Kherrata.

Ce site est localisé entre la route nationale numéro neuf et le flanc sud de la rive gauche de l'Oued Agrioun et il se trouve entre le carrefour du viaduc Bordj Mira soit au PK0+00m et à l'entrée de la ville de Kherrata au PK7+650m.

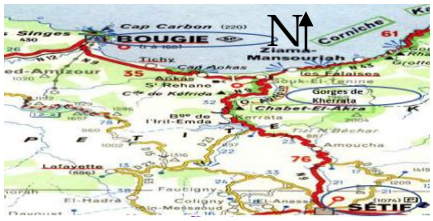


Figure 1. Situation géographique des gorges de Kherrata d'après Carte Michelin, N°172,1958, Algérie, Tunisie

Le secteur des gorges de Kherrata est caractérisé par un substratum calcaire et dolomitique, structuré en lames épaisses, dont certaines sont plissées. Il est affecté par des failles avec ou sans brèches de faille, de diaclases en réseaux et de fractures nombreuses ouvertes ou fermées. Des recouvrements d'éboulis, travertins ou brèches de pente à blocs, masquent localement les calcaires et les dolomies.

A. La description du talus rocheux au PK1+172m

Le talus rocheux, constitué par des marnes schisteuses à moyen pendage vers l'ouest et dont la direction des couches varie (S_0: N30°/30°W, N10°/30°W) et recouvertes par une épaisse brèche de pente à blocs, est affecté par des mouvements localisés de glissement. La Fig.2 présente une coupe longitudinale du talus rocheux au PK1+172m.

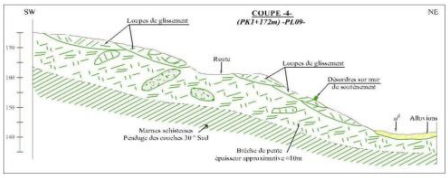


Figure 2. Coupe transversale du talus rocheux de PK1+172m

VI. EXEMPLE : ESTIMATION DE LA STABILITE DE TALUS ROCHEUX

B. Détermination du facteur de sécurité par l'analyse déterministe

L'analyse déterministe est effectué par :

Le logiciel Phase 2D et Slide 2D, afin de déterminer un facteur de sécurité unique. Les modèles géométriques utilisés pour la modélisation sont présentés dans la Fig.3.

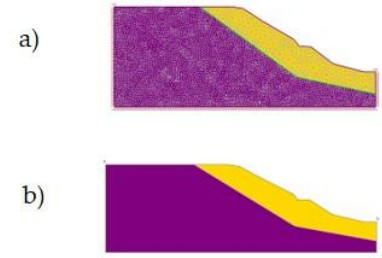


Figure 3. Modèles géométriques utilisés pour la modélisation, a) modèle de Phase 2D, b) modèle de Slope 2D

Les caractéristiques géotechniques des roches qui constituent le talus rocheux de PK1+172 sont résumées dans le tab.1. Les valeurs de ces propriétés des roches sont utilisées pour calculer la valeur du facteur de sécurité.

Tableau 1. Les paramètres des couches de roches qui constituent le talus rocheux talus rocheux

Type des roches	*Paramètres géotechniques de la roche*
Marnes Schisteuses	γ *=21KN/m³;c' =7360KPa;E=12700KPa*
Bréche de pente	γ *=22KN/m³;* c' *=5KPa;* φ' *=35°; E=72000KPa*

Tableau 2. Les paramètres statistiques de la marne schisteuse

Paramètres statistiques de la marne schisteuse		
Cohesion(Kpa)	*Poids volumique(KN/m³)*	*Module de Young(Kpa)*
Distribution:Normal	Distribution:Normal	Distribution:Normal
Moyenne:7360	Moyenne:21	Moyenne: 127000
Écart type:1967	Écart type:0.8	Écart type:58000
Max:5901	Max:2.4	Max:174000
Min:5901	Min:2.4	Min: 174000

Tableau 3. Les paramètres statistiques de la brèche de pente

Paramètres statistiques de la bréche de pente		
Cohesion(Kpa)	*Poids volumique(KN/m³)*	*Module de Young(Kpa)*
Distribution:Normal	Distribution:Normal	Distribution:Normal
Distribution:Normal	Moyenne:22	Moyenne: 72000
Moyenne:5	Écart type:0.5	Écart type:3000
Écart type:3.2	Max:1.5	Max:9000
Max:9.6	Min:1.5	Min:9000

B. Estimation de la stabilité de talus rocheux par la méthode probabiliste

Les valeurs moyennes et statistiques des paramètres géotechniques des roches pour la détermination du facteur de sécurité moyen, de son écart type, de la fiabilité et de la probabilité de rupture, sont consignées dans les Tab.1,2 et 3.

VII. Resultats et Discussions

La méthode de Bishop a donné un facteur de sécurité d'ordre de 1.256. La méthode d'éléments finis a calculé un facteur d'environ de 1.29 qui est supérieur à celui déterminé par Bishop de 3%. Ces deux méthodes déterministes ont estimé des facteurs de sécurité inférieurs à la valeur critique de 1.5 en condition statique. Les facteurs de sécurité obtenus par les deux méthodes déterministes traduisent l'instabilité du talus rocheux des gorges de Kherrata.

Les Fig.4 et 5 représentent les valeurs de facteur sécurité du talus rocheux au PK1+172.

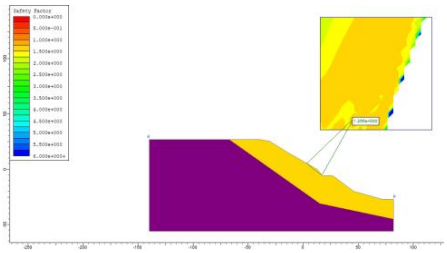


Figure 4. Facteur de sécurité déterminé par la méthode simplifiée de Bishop

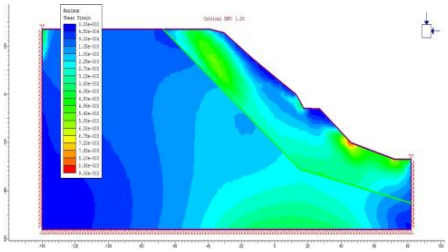


Figure 5. Facteur de sécurité déterminé par la méthode des éléments finis

Les résultats de l'analyse probabiliste sont consignés dans le tab.4. Les facteurs de sécurité, qui sont déterminés par les deux méthodes probabilistes, la méthode d'approximation ponctuelle de Rosenblueth et la simulation de Monte-Carlo, sont inférieurs à 1.5 : ils sont de 1.24 et 1.11 respectivement. Ceci confirme l'instabilité du talus rocheux au PK 1+172.

Tableau 4. Les résultats de l'analyse probabiliste

Les méthodes probabilistes	***Simulations de Monte-Carlo***	***Méthode d'approximation ponctuelle de Rosenblueth***
Les résultats	FS_{moyen}=1.11 β =*1.88* PF =*3.3%*	FS_{moyen}=1.24 $\sigma_{FSmoyen}$=0.17 PF= 7.45%

L'écart entre la méthode de Bishop et la simulation de Monte-Carlo est d'environ 18 %. L'écart est de 3 % entre la méthode d'approximation ponctuelle de Rosenblueth et celle des éléments finis.

La probabilité de rupture de la simulation de Monte-Carlo est autour de 3.3%, estimé à 7.45 % par la méthode d'approximation ponctuelle de Rosenblueth avec une différence de 4% de la simulation de Monte-Carlo.

Les résultats des calculs déterministes et probabilistes se convergent avec des écarts moyens. Malgré ces écarts, les résultats des méthodes déterministes traduisent l'état instable de ce talus rocheux.

La fiabilité prédite par la méthode de Monte-Carlo combinée à la méthode simplifiée de Bishop est d'environs 1.88. La performance de cette fiabilité est insatisfaisante à pauvre [12].

VIII. Conclusion

D'après les résultats obtenus, nous pouvons conclure que :

1. L'approche déterministe donne généralement des valeurs conservatrices du facteur de sécurité puisque les paramètres d'entrées ont une valeur unique et la variation spatiale de ces paramètres n'a pas été prise en compte.

2. L'approche probabiliste donne une valeur de facteur de sécurité moyenne puisque les valeurs des paramètres géotechniques de talus rocheux sont traduites par des paramètres statistiques (moyenne, écart-type, maximum, minimum) et une distribution probabiliste.

3. L'étude des éléments finis a confirmé les résultats de la méthode de Bishop. Selon les résultats de l'approche déterministe le talus rocheux étudié est instable.

4. Les écarts entre les méthodes probabilistes et déterministes d'ordre de 3% et 18% justifient la performance insatisfaisante à pauvre de la fiabilité des résultats des méthodes déterministes mais qui n'a pas un grand effet sur les valeurs des facteurs de sécurité.

5. Les valeurs des facteurs de sécurité déterminées par les méthodes déterministes et probabilistes présentent l'état instable du talus rocheux des gorges de Kherrata.

References

[1] M. Abdulai and M. Sharifzadeh,"*Uncertainty and reliability analysis of open pit rock slopes: a critical review of methods of analysis,*"Geotechnical and Geological Engineering, 37,1223-1247,2019.

[2] M. Abbaszadeh et al. , "*Uncertainty and reliability analysis applied to slope stability: a case study from Sungun copper*

mine," Geotechnical and Geological Engineering 29, 581-596, 2011.

[3] E. Eberhart, "*Rock slope stability analysis, utilization of advanced numerical techniques*,"Technical Report 41, UB C – Vancouver, Canada, 2003.

[4] D. Mendiel. "*Analuse inverse dans le calcul geotechnique-application au calcul de la stabilite des talus,*" thése de doctorat,Université de Annaba,2012.

[5] A.W. Bishop, "*The Use of The Slip Circle in the Stability Analysis of Earth Slope*," Geotetoratchnique, 5(1), 7-17,1955.

[6] T.K. Mebrahtu et al., "*Slope stability analusis of deep-seated landslides using limit eauilibrium and finite element methods in Debre Sina area, Ethiopia*," Bulletin of Engineering Geology and the Environment, 81(10), 403,2022.

[7] D.V. Griffiths and P.A. Lane,"*Slope stability analysis by finite elements*,"*Geotechnique*,49:387–403,1999.

[8] C.K. Aswathi et al., "*Stabilitu assessment of reinforced rock slope based on two-dimensional finite element approach: A Himalayan case study*," Geotechnics for Transportation Infrastructure: Recent Developments, Upcoming Technologies and New Concepts, Volume 1. Springer Singapore, 2019.

[9] T. Matsui and K.C. San, "*Finite element slope stabilitu analusis bu shear strength reduction technique*, "*Soils and foundations*, *32*(1), 59-70,1992.

[10] R.Hammah et al., "*Numerical modelling of slope uncertaintu due to rock mass jointing*," *In : Proceedings of the international conference on rock joints and jointed rock masses,2009.*

[11] E. Rosenblueth, "*Point estimates for probabilitu moments*, "Proceedings of the National Academy of Sciences, 72(10), 3812-3814,1975.

[12] USAC, "*Risk-based analusis in geotechnical engineering for support of planning studies, engineering and design*,"1997.

Effets de la réparation des joints de soudure sur la fiabilité des pipelines enterrés soumis à la corrosion

Sahraoui Yacine
Laboratoire Électromécanique et Sûreté de Fonctionnement, Université Mohamed-Chérif Messaadia Souk Ahras, BP 1553, Souk Ahras, 41000, Algérie
y.sahraoui@univ-soukahras.dz

Nahal Mourad
Laboratoire Électromécanique et Sûreté de Fonctionnement, Université Mohamed-Chérif Messaadia Souk Ahras, BP 1553, Souk Ahras, 41000, Algérie
m.nahal@univ-soukahras.dz

Alaa Chateauneuf
Université Clermont Auvergne, CNRS, SIGMA Clermont, Institut Pascal, F-63000 Clermont-Ferrand, France
alaa.chateauneuf@cideco.tech

Résumé—
Le présent travail porte sur une étude des effets de la réparation des joints de soudure sur la fiabilité des pipelines corrodés en tenant compte de la variabilité spatiale de la dureté et l'agressivité du sol. Tout d'abord, une étude expérimentale a été réalisée sur des tubes testés de type X70, ayant subies une série de réparations successives. Les objectifs de cette partie sont de mesurer la dureté après chaque réparation au niveau des différentes sous-zones de soudure (i.e. Métal de Base "M.B.", Zone Affectée Thermiquement "Z.A.T." et Métal d'Apport "M.A.") et de dériver la corrélation statistique « limite élastique-dureté ». Ensuite, une méthode probabiliste est présentée pour évaluer la fiabilité du système en fonction du temps des tuyaux enterrés soumis à la corrosion, en tenant compte de la variabilité spatiale de la dureté et l'agressivité du solsur toute la longueur du tuyau.Les probabilités de défaillance au niveau des différentes sous-zones de soudure sont calculées par la méthode de Monte-Carlo et la décomposition de Karhunen-Loève a été utilisée pour modéliser cette variabilité spatiale de la dureté et l'agressivité du sol.Les résultats montrentquel'effet des réparations de joints de soudure sur la fiabilité du système est notable, en particulier de la 1ère réparation.

Mots-Clés—
Joints soudés, Fiabilité de pipelines, Variabilité spatiale

I. INTRODUCTION

Les pipelines tiennent une place importante dans les industries gazière et pétrolière, et contribuent pleinement à son développement durable en assurant leur fonction de production, de transport et de distribution [1]. Les pipelines sont des canalisations le plus souvent en acier, constituées de segments linéaires soudés. La qualité des soudures n'est malheureusement pas toujours celle que l'on attend, les normes définissent les critères d'acceptabilité des défauts présents dans une soudure. Cependant, des paramètres technico-économiques obligent les concepteurs et les gestionnaires à éliminer localement les défauts issus de la fabrication par des réparations, afin d'éviter la coupe ou le rejet total du tronçon. Même si la réparation de la soudure fait partie du processus de fabrication des pipelines, elle peut causer une variabilité des propriétés mécaniques au niveau du cordon de soudure en raison de l'apport considérable de température à travers les cycles thermiques de soudage.

L'acier présente l'inconvénient d'être sensible à la corrosion, notamment lorsqu'il est enterré dans le sol [2,3]. Ce dernier est un amalgame complexe de solides, liquides et gaz, dont l'agressivité est très variable et peut parfois conduire à des corrosions extrêmement rapides. Toutefois, la plupart des modèles prédictifs actuels des pipelines enterrés ne tiennent pas compte de la forte variabilité spatiale des propriétés mécaniques et de la corrosion [4,5,6].

Dans ce contexte, le présent travail porte sur l'étude des effets de la multiple réparation des joints de soudure sur la fiabilité des pipelines corrodés en tenant compte de la variabilité spatiale de la dureté et l'agressivité du sol. Tout d'abord, nous avons réalisé une étude expérimentale pour mesurer la dureté [7] après chaque réparation au niveau des différentes sous-zones de soudure (i.e. Métal de Base "M.B.", Zone Affectée Thermiquement "Z.A.T." et Métal d'Apport "M.A."). Ensuite, la décomposition de Karhunen-Loève a été utilisée pour modéliser la variabilité spatiale de la dureté et l'agressivité du sol, en tenant compte de l'auto-corrélation [8]. Dans ce travail, le pipeline est modélisé par un système en série. Les probabilités de défaillance du système sont calculées par la méthode de Monte-Carlo.

II. PROCEDURES EXPERIMENTALES

L'étude de l'effet de la réparation s'est effectuée sur la base de mesure de la dureté des différentes zones du joint de soudure après chaque réparation [9]. Pour ce faire, une plaque a été découpée à partir d'un pipeline soudé en spiral de diamètre de 42 pouce et d'épaisseur d'environ 12.9 mm. La découpe a été faite de façon à avoir le joint de soudure le long de la plaque. Avant toute réparation, des échantillons destinés à la mesure de la dureté de la soudure seine ont été découpés. Le reste de la plaque a subi la première réparation le long du joint de soudure, On procède au prélèvement d'échantillons correspondant à cette opération. Le reste de la plaque a subi la deuxième réparation intégralement et des échantillons de cette étape ont été découpés. On procède de la sorte pour la troisième réparation.

La réparation de la soudure est une opération manuelle qui fait partie du procédé de fabrication et se réalise selon les étapes suivantes :

1. Le gougeage : Cette étape correspond à un enlèvement de la matière du cordon de soudure en creusent un sillon le long du joint. Ainsi, dans le cas d'une présence d'un défaut, il sera éliminé avec le métal éjecté. Cette opération est effectuée au moyen d'un dispositif appelé Arc-air.
2. Le remplissage : Une première couche de métal d'apport est déposée au moyen d'électrode, dans le but de réaliser un premier remplissage du creux obtenu par gougeage.
3. Le remplissage-finition : On obtient par remplissage l'épaisseur finale du bourrelet.

La mesure de la dureté est effectuée conformément à la norme API 5L selon le schéma de disposition de 16 points de mesure comme représenté sur la Figure 1.

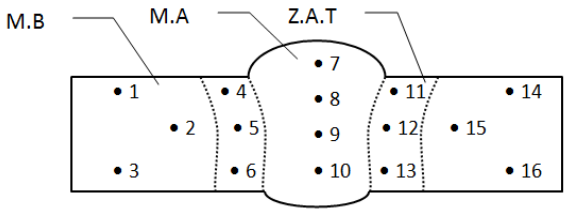


Figure 1. Schèma de disposition des points de mesures de la dureté

Le choix de la mesure de dureté comme moyen de relever la variabilité des propriétés mécaniques est justifié, étant donné qu'il est possible à travers cet essai d'avoir des mesures locales relatives à chaque zone et pour chaque opération de réparation. Les mesures ont été effectuées au niveau du laboratoire de contrôle qualité de l'entreprise ALFAPIPE Annaba. Le dispositif expérimental utilisé est du type ZWICK HV10, Figure 2, qui permet d'obtenir la dureté en HV10.

Figure 2. Dispositif de mesure de la dureté HV10

Afin d'intégrer les mesures de dureté dans le modèle fiabiliste, il a été nécessaire d'effectuer la conversion de ces dernières en contraintes. Pour se faire, la méthodologie développée dans les travaux de S. H. HASHEMI [7] a été adoptée. Le principe de cette méthode repose principalement sur une observation relative à la dispersion des mesures de dureté et des contraintes. Il a été constaté que la dureté se caractérise par une dispersion moins importante par rapport à la contrainte pour un échantillonnage donné. En conséquence, chaque mesure de dureté regroupe un ensemble de valeurs de contrainte. La relation entre ces deux grandeurs est établie entre la moyenne des contraintes enregistrées pour chaque mesure de dureté et la valeur même de cette mesure de dureté. Ainsi la corrélation est obtenue suite à un lissage linéaire du nuage de points correspondant au couples (HV10, contrainte). L'auteur a démontré que la relation entre la dureté et la contrainte (limite élastique ou la résistance à la traction) est linéaire. Aussi, l'auteur a mis en évidence que la relation entre la limite élastique et la résistance à la rupture est également linéaire.

Dans le cas de l'acier X70, la conversion de la dureté HV10 en limites élastiques (fy) dans le cas du métal de base est obtenue au moyen de la relation suivante :

$$f_y = 0.675 \cdot Hv + 38 \qquad (1)$$

La Figure 3, montre une représentation graphique de la corrélation linéaire entre les mesures de dureté HV10 et la limite élastique correspondantes aux différentes zones.

Pour introduire le joint soudé, différentes approches sont considérées dans cette étude. Une première approche peut être adoptée sous l'hypothèse que le joint soudé fait partie du continuum du métal de base. Cela permet en conséquence de considérer que le joint

soudé obéit à la même corrélation précédemment établie à partir du métal de base. Ainsi, il serait possible de quantifier la limite élastique du joint, à partir de son relevé de mesures de dureté. On suppose également que cette relation reste valable pour les différentes réparations.

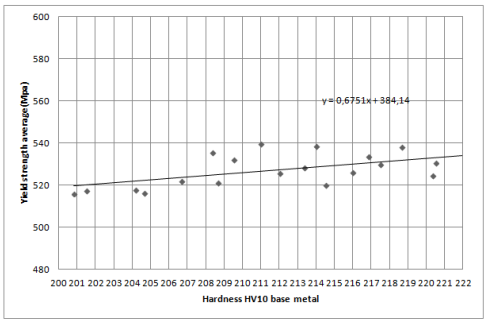


Figure 3. Corrélation linéaire entre la limite d'élasticité et la dureté de l'acier X70 (cas du métal de base)

Une deuxième approche peut être considérée du fait que la contrainte au niveau du joint de soudure est différente de celle du métal de base. En effet, les essais de traction sur des éprouvettes comportant un joint de soudure présente une moyenne des contraintes de rupture plus élevée relativement à celle du métal de base. Cependant, on ne dispose que des valeurs expérimentales de la résistance à la traction et de la dureté au niveau du joint. En fonction de ces deux paramètres, une nouvelle corrélation peut être établie et qui, à priori, peut refléter mieux l'aspect relationnel entre les propriétés mécaniques propres à cette partie de pipeline. Cette approche est développée en tenant compte des trois zones (métal de base, Z.A.T. et métal d'apport), sous les hypothèses suivantes :

- La relation entre la résistance à la rupture et la dureté HV10 est linéaire
- La relation entre la limite élastique et la résistance à la rupture est linéaire
- Les relations précédentes restent valables dans le cas des réparations

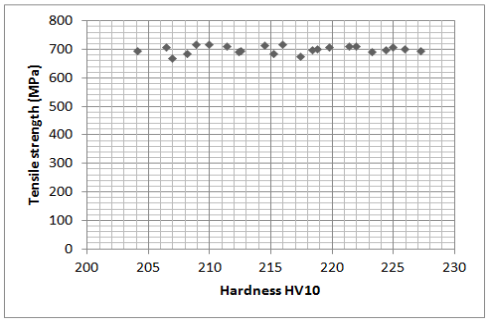


Figure 4. Corrélation entre la contrainte de rupture et la duretés HV10

Le lissage du nuage de points résultants entre les contraintes de rupture et la dureté dans le joint de soudure a abouti à une représentation linéaire, Figure 4, où :

$$\sigma_r = 0.1973 \cdot Hv + 658 \qquad (2)$$

L'obtention de la limite élastique peut s'effectuée à partir de la relation entre la limite élastique et la contrainte de rupture par la relation:

$$f_v = 1.002\sigma_r - 11 \qquad (3)$$

La dernière approche est fondée sur l'observation faite sur les mesures de dureté après chaque réparation. Sachons que l'action d'enlèvement et d'ajout du métal suite à une réparation se localise sur la portion supérieure du métal d'apport, qui se définie par les points de mesure 7 et 8, Figure 1. Il serait intéressant de se confiner au niveau de cette zone, en ne tenant compte que des mesures de dureté des point 7, 8, 9 et 10 qui jouent un rôle prépondérant dans l'augmentation de la dispersion d'une réparation à une autre. Ces valeurs sont exploitées dans la corrélation relative au joint soudé, ce qui va mieux révéler la nature dispersive et l'effet des mesures au niveau du joint soudé.

III. DEFINITION DU SYSTEME

On désigne par « pipelines », les oléoducs ou gazoducs qui sont des canalisations le plus souvent en acier, généralement enterrées et reposent directement sur le sol, dans lesquelles transitent, sous des pressions relativement élevées, des produits pétroliers, liquides ou gazeux, sur de longues distances. L'avantage de l'utilisation de ces constructions en acier est principalement la réduction du coût global du projet. Les tuyaux soudés doivent supporter la pression interne du fluide et les conditions défavorables de l'environnement externe.

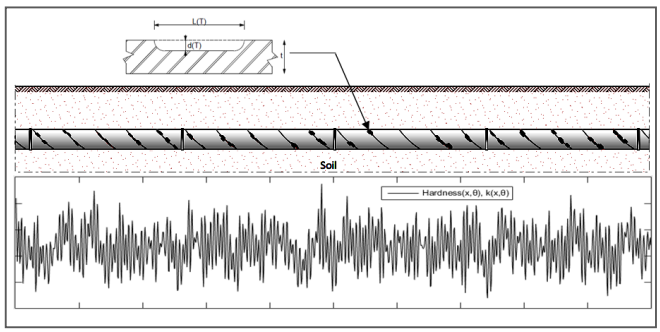


Figure 5. Pipeline formé de plusieurs segments

Si l'on considère que le pipeline est en contact direct avec le sol et que le cordon de soudure est une zone singulière [10], la variabilité spatiale des propriétés

mécaniques de cette zone et de l'agressivité du sol doit être prise en compte, puisqu'elle affecte la fiabilité globale du système. Dans notre étude, le pipeline est modélisé par un système en série composé de plusieurs segments et présentant des défauts de corrosion localisée à plusieurs endroits (Figure 5). Le mode de défaillance correspond à la tenue du pipeline sous l'effet de la corrosion et de la pression interne.

A. Modèle probabiliste de dégradation

La défaillance des tuyaux est identifiée à l'aide d'un modèle semi-empirique basé sur la mécanique de la rupture, pour déterminer la pression à laquelle la conduite devient défaillante, en fonction de la géométrie du défaut de corrosion. La pression de rupture est exprimée par [11]:

$$P_r = 2(f_y + 68.95)\left(\frac{t}{D}\right)\left[\frac{1-\frac{d(T)}{t}}{1-\frac{d(T)}{tM}}\right] \quad (4)$$

Dans cette expression, *t* est l'épaisseur du tuyau, *D* est son diamètre, T est l'âge de la conduite, fy est la limite élastique du cordon de soudure, et M le facteur de Folias, aussi connu comme facteur de bombement, est un facteur semi-empirique qui couvre les aspects de la mécanique de la rupture (i.e. une majoration de la contrainte). En adoptant des champs stochastiques, la limite élastique, la profondeur d(T) et la longueur L(T) d'un défaut longitudinal de corrosion en fonction du temps, peuvent être déterminées par:

$$f_y = \begin{cases} 0.675 \cdot Hv(\vec{x},\theta) + 384.1\,, & 1^{ère}\text{Approche} \\ 0.198 \cdot Hv(\vec{x},\theta) + 545.1, & 2^{ème}\text{Approche} \end{cases} \quad (5)$$

$$d(T) = k(\vec{x},\theta) \quad (6)$$

$$L(T) = \gamma\, k(\vec{x},\theta) \quad (7)$$

Où est la dureté mesurée du cordon de soudure, et sont des paramètres incertains de la corrosion, qui peuvent être dérivés à partir des propriétés du sol, et sont respectivement les variables spatiales et stochastiques, et est le rapport entre la longueur et la profondeur de la corrosion localisée. Les grandeurs et définissent le mécanisme de corrosion et sont identifiées par des études statistiques, en fonction de la concentration d'agents agressifs dans le sol (i.e. chlorure, sodium, etc.). La variable x correspond à l'abscisse longitudinale du pipeline, et est une variable générique représentant l'aléa. La fonction d'état limite G_{ij}, à chaque défaut individuel j dans le segment i, correspond à la marge de sûreté, qui est classiquement définie par la différence entre la résistance de la conduite et la pression appliquée à la même section du tuyau :

$$G_{ij} = P_{rij} - P_a \quad (8)$$

Par conséquent, la probabilité de défaillance individuelle (i.e. d'une section) est la suivante :

$$P_{fij} = P[G_{ij} \le 0] \quad (9)$$

B. Evaluation des probabilités de défaillance

1) Champ stochastique de corrosion.

Parmi les outils performants de génération de réalisations de champs corrélés, la décomposition de Karhunen-Loève permet d'approximer les champs aléatoires *K* et *Hv* par une série finie de variables aléatoires indépendantes normalisées :

$$k(\vec{x},\theta) = \bar{k}(\vec{x}) + \sum_{i=1}^{M} \sqrt{\lambda_i}\,\phi_i(\vec{x})\xi_i(\theta) \quad (10)$$

$$Hv(\vec{x},\theta) = \overline{Hv}(\vec{x}) + \sum_{i=1}^{M} \sqrt{\lambda'_i}\,\phi'_i(\vec{x})\xi_i(\theta) \quad (11)$$

Un des principaux atouts de cette méthode est qu'elle permet le découplage des variables spatiales $\vec{x}$ et stochastiques θ. Les premiers termes de la décomposition sont les espérances spatiales des champs aléatoires $\bar{k}(\vec{x})$, $\overline{Hv}(\vec{x})$. La dépendance spatiale apparaît dans les modes propres (λ_i, ϕ_i), (λ'_i, ϕ'_i), du noyau de covariance, où λ_i, λ'_i sont les valeurs propres et ϕ_i, ϕ'_i sont les vecteurs propres correspondants.

2) Probabilités de défaillance pour le système en série

En raison de la complexité du système spatio-temporel dans notre étude, la méthode de Monte-Carlo (MC) est utilisée pour le calcul des probabilités de défaillance et de réparation. La probabilité de défaillance d'un seul segment i présentant un nombre de défauts de corrosion est donnée par :

$$P_{F_{Segment}} = P\left(\bigcup_{j=1}^{m_i}[G_{ij} \le 0]\right) \quad (12)$$

En conséquence, la probabilité de défaillance d'un système composé de n segments peut s'écrire sous la forme :

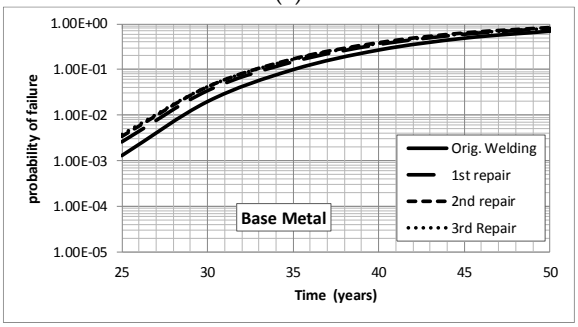
$$P_{F_{Système}} = P\left[\bigcup_{i}^{n}\left(\bigcup_{j=1}^{m_i}[G_{ij} \leq 0]\right)\right] \quad (13)$$

Dans cette expression, le pipeline est considéré comme un système en série composé de sous-systèmes en série. En effet,la canalisation est une longue conduite tubulaire composée d'un certain nombre de segments soudés et chaque segment présente un certain nombre de défauts de corrosion localisés. La défaillance à tout défaut implique la défaillance du segment, et de même la défaillance de tout segment entraîne la défaillance de l'ensemble du système.

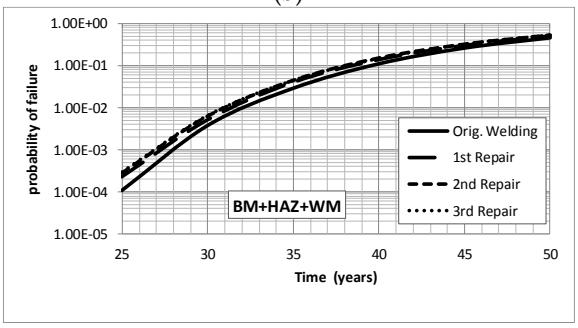
IV. IMPACT DE LA REPARATION DES JOINTS DE SOUDURE SUR LA FIABILITE DES PIPELINES

L'application numérique aux pipelines corrodés vise à montrer l'effet de la réparation des joints de soudure sur la fiabilité globale des pipelines en tenant compte de la variabilité spatiale de l'agressivité du sol et de la dureté après chaque réparation au niveau des différentes sous-zones de soudure. Le pipeline testé est de type X70, avec le diamètre D=42 pouce et l'épaisseur de paroi t=12.9 mm, et soumis à la pression interne P=8.5 MPa, où toutes les variables sont distribuées selon la loi log-normale. Dans ce travail, nous considérons un pipeline d'une longueur de 200 km,enterré dans un sol de type « générique » contenant plusieurs types de sol, où les valeurs moyennes de k et n sont 0,164 et 0,780, respectivement [8,12].

(a)



(b)



(c)

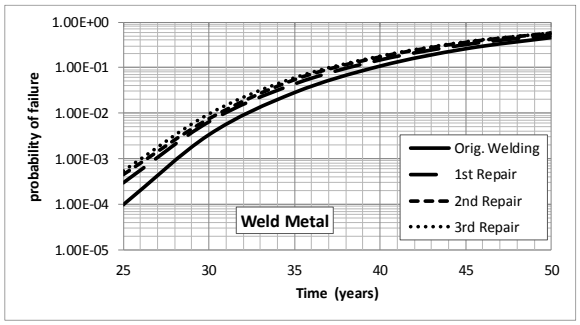


Figure 6. Effet des réparations successives

Tableau 1. Valeurs de dureté Hv des tubes testés de type X70

Zones	*Soudures*	*Moyennes (MPa)*	*C.V*
MB+ZAT+MA	*S. Originale*	*217.56*	*0.033*
	1ère réparation	*210.56*	*0.06*
	2ème réparation	*205.5*	*0.065*
	3ème réparation	*208.13*	*0.07*
MA	*S. Originale*	*221*	*0.03*
	1ère réparation	*201*	*0.08*
	2ème réparation	*198*	*0.1*
	3ème réparation	*199*	*0.12*

Afin d'analyser l'influence des réparations successives sur la fiabilité des pipelines corrodés, trois réparations sont considérées. Les valeurs de la dureté après chaque réparation au niveau des différentes sous-zones de soudure sont indiquées dans le Tab. 1.

L'effet des réparations successives est maintenant considéré dans l'analyse de fiabilité des pipelines corrodés. En adoptant ici des champs stochastiques avec des longueurs de corrélation de 50m et 5m correspondant respectivement à la longueur de corrélation spatiale de l'agressivité du sol et des propriétés mécaniques du cordon de soudure. Les Figures 6.a-c montrent l'évolution temporelle de la probabilité de défaillance des différentes zones du joint de soudure après chaque réparation ainsi que la probabilité de défaillance du système après la soudure originale.

La Figure 6.a représente une première approche adoptée sous l'hypothèse que le joint soudé fait partie du continuum du métal de base. Les Figures 6.b et 6.c représentent la deuxième approche considérée du fait que la contrainte au niveau du joint de soudure est différente de celle du métal de base, où la Figure 6.b considère toutes les zones (MB+ZAT+MA), tandis que la Figure 6.c considère uniquement le métal d'apport

(MA).Ces courbes fournissent une image claire du niveau du risque nominal de défaillance du tube pour différentes zones du joint de soudure. Comme prévu, le métal de base est moins fiable que le cordon de soudure. Par exemple, après 25 ans de service, la probabilité de défaillance du système est estimée à 10-4 considérant le joint de soudure, par contre la probabilité de défaillance du système est d'environ 10-3 considérant uniquement le métal de base. En observant les Figures 6.a-c, nous pouvons voir l'effet des réparations successives sur la fiabilité des pipelines corrodés des différentes sous-zones de soudure. Dans tous les cas, les réparations tendent à réduire la fiabilité du système, notamment la 1ère réparation. Ceci peut être expliqué par la variabilité des propriétés mécaniques au niveau du cordon de soudure en raison de l'apport considérable de température à travers les cycles thermiques de soudage, où cette variabilité présente une certaine stabilité après la 1ère réparation. En comparant les Figures 6a-c, nous pouvons constater aussi que l'effet des réparations est plus notable dans le cas du métal d'apport (Figure 6.c) par rapport aux autres cas, qui peut être compris par l'augmentation de la dispersion des mesures due à l'action d'enlèvement et d'ajout du métal suite à une réparation se localise sur la portion supérieure du métal d'apport.

V. CONCLUSION

Les effets de la réparation des joints de soudure ont été étudiés afin d'évaluer la fiabilité des pipelines enterrés de type X70 en tenant compte de la variabilité spatiale de la dureté et l'agressivité du sol. Les simulations de Monte Carlo ont permis de calculer les probabilités de défaillance du système. Dans le cadre de cette étude, la partie expérimentale a révélée que la réparation n'altère pas de manière significative les moyennes des mesures. Cependant, l'effet notable reste par rapport à la dispersion et au coefficient de variation de la dureté. Comme conclusion, à long terme et en présence de la corrosion, la réparation peut avoir un effet sur la fiabilité du système et notamment celui de la 1ère réparation. Cela, au regard des résultats obtenus et de la procédure de conversion des duretés réalisée dans le cadre de cette étude. En perspectives de ce travail, les formulations présentées ici peuvent être utilisées dans l'élaboration d'une stratégie optimale de maintenance et d'inspection des joints de soudure.

REFERENCES

[1] H.A. Kishawy, H.A. Gabbar, "Review of pipeline integrity management practices," International Journal of Pressure Vessels and Piping, 2010, vol. 87, pp. 373–380.

[2] I.S. Cole, D. Marney, "The science of pipe corrosion: A review of the literature on the corrosion of ferrous metals in soils," Corrosion Science, 2012, vol. 56, pp. 5–16.

[3] J.L. Alamilla, M.A. Espinosa-Medina, E. Sosa, "Modelling steel corrosion damage in soil environment," Corrosion Science, 2009, vol. 51, pp. 2628–2638.

[4] M. Ahammed, R.E. Melchers, "Probabilistic analysis of underground pipelines subject to combined stresses and corrosion," Engineering Structures, 1997, 19, pp. 988–994.

[5] A. Amirat, A. Chateauneuf, K. Chaoui, "Reliability assessment of underground pipelines under the combined effect of active corrosion and residual stress," International Journal of Pressure Vessels and Piping, 2006, vol. 83, pp. 107–117.

[6] Y. Sahraoui, R. Khelif, A. Chateauneuf, "Maintenance planning under imperfect inspections ofcorroded pipelines," International Journal of Pressure Vessels and Piping, 2013, vol. 104, pp. 76-82.

[7] S.H. Hashemi, "Strength–hardness statistical correlation in API X65 steel," Materials Science and Engineering A, 2011, vol. 528, pp. 1648–1655.

[8] Y. Sahraoui, A. Chateauneuf, "The effects of spatial variability of the aggressiveness of soil on system reliability of corroding underground pipelines," International Journal of Pressure Vessels and Piping, 2016, vol. 146, pp. 188–197.

[9] Y. Sahraoui, M. Benamira, M. Nahal, F. Nouadria, A. Chateauneuf, "The effect of welded joint repair on a corroded pipeline reliability subjected to the hardness spatial variability and soil aggressiveness," Engineering Failure Analysis, 2020, vol. 118, pp. 104854.

[10] M. Nahal, A. Chateauneuf, Y. Sahraoui, "Reliability analysis of irregular zones in pipelines under both effects of corrosion and residual stress," Engineering Failure Analysis, 2019, vol. 98, pp. 177–188.

[11] M. Ahammed, "Probabilistic estimation of remaining life of a pipeline in the presence of active corrosion defects," International Journal of Pressure Vessels and Piping, 1998, vol. 75, pp. 321–329.

[12] F. Caleyo, J.C. Vel_azquez, A. Valor, J.M. Hallen, "Probability distribution of pitting corrosion depth and rate in underground pipelines: a Monte Carlo study," 2009, Corrosion Science, vol. 51, pp. 1925–34.

III

FIABILITÉ ET SÉCURITÉ DES SYSTÈMES INFORMATIQUES

Sommaire

Quality of service in Web services system

Bernine Nassima
LaMOS Research Unit,
Operational Research Department,
University of Bejaia, Algeria
nassima.bernine@univ-bejaia.dz

Aissani Djamil
LaMOS Research Unit,
Operational Research Department,
University of Bejaia, Algeria
djamil.aissani@univ-bejaia.dz

***Abstract*—**
The Web services are accessible applications on Internet realize each one a specific task. To provide a solution to a complex task, we can gather Web services to form another one of them; So, we talk about a composition of Web services. This paper defines a model to evaluate the performance of Web services system. We have modeled with Petri nets stochastic. This approach allowed us to have an analytical model of Web services system, which takes into account two different types of arrivals: clients who demand a service and Web services who provide the service. The performances of this model are calculated using a simulator GRIF. Analysis results are presented to show their impact in virtual Web services; simple or composite. Furthermore, we compute the system's response time, and the Average number of clients in the system ,in terms of the arrival rate of clients requests.

***Keywords*—**
Web Services Composition, Petri nets model, Performance evaluation, Modeling, Simulator GRIF.

I. Introduction

Web services are self-describing, modular, loosely coupled applications that provide a standards-based model for programming and deploying applications that run through the web infrastructure [1]. The discovery of Web services consists in finding the appropriate mappings between the elements of user needs and the elements of existing Web services. These needs can be covered by a simple Web service, or a complex Web service resulting from the composition task [2], [3]. To provide a solution to a complex task, we can group Web services to form a single one; we then speak of the composition of Web services, and we get a composite Web service (Nacer and Aissani, 2014; Yugen and Yonggang, 2010; Nacer and Djebari et al., 2017). This latter does not belong to any supplier, because the Web services in several components are produced by the different independent suppliers (Nacer and Aissani, 2014; Nacer and Djebari et al., 2017); The objective of Web services is to facilitate access to applications between companies and thus simplifying data exchanges. They're chasing an old computer dream distributed network where applications could interoperate across the network, independently their platform and implementation language.
In Web service composition approach implementing SOA (Simple Object Access) architecture, The three (3) entities (Provider, Client, UDDI (Universal Description Discovery and Integration)) interact as follows Fig. 1:

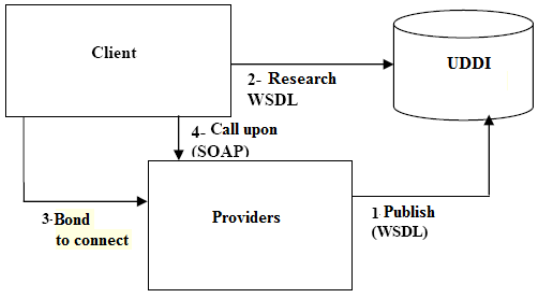


Figure 1. Web services architecture.

- The service provider defines the description of its service, and publishes it in a universal service directory of Web services, UDDI (Universal Description Discovery and Integration);
- The client uses the search facilities available in the universal directory to find and select a given service;
- The client then examines the description of the selected service to retrieve the necessary information that allows it to connect to the service provider and interact with the implementation of the service.

II. Related Work

During the last few years, the problem of Web services performance evaluation has received a lot of attention by many researchers, in order to improve the customers satisfaction.
In [4], the authors decomposed the requests in sub-Queries to different elementary Web services, and then were merged into a final result. Thy assumed that an elementary Web service can be invoked with a constant probability. They broke up the requests into "n" elementary services, where each elementary services will call upon "n" Web services. They calculated the response

time of clients in terms of service rate. In [5], The authors presented the composed services with different performance parameters. In order to resolve the composed services, evaluate its performance metrics using Sub-Optimal and Heuristic approaches, and implemented it in java or C sharp language. They calculated mean: Response time, turnaround time. In [6], The author proposed a new approach for efficient service composition based on abstraction refinement. Instead of considering individual services during composition, they proposed several abstractions to form service groups and the composition was done on these abstract services, using an algorithm for an abstraction procedure implemented in Java. They calculated the number of abstract and average composition time.

In [7], the authors propose a model to quantify availability and reliability of atomic services using Markov Chain model and Weibull analysis respectively. Markov Chain models are used to study systems that could be represented as discrete states.

In [8], under the same hypotheses as in [04], the authors use another method (Stochstic Automata Networks) to calculate the average response time, by taking into account the number of Web services with respect to whether they are: constant or variable of services.

In [9], the authors have presented an optimization approach for the composition of Web services for systems Service-Oriented Architecture (SOA) on a large scale, which is the subject of the constraints of quality of service (QoS). In particular, the authors influence on the composition model for the flexible specification of QoS constraints by employing hierarchical constraints. In [10], the authors proposed a reliability calculation method for Web service compositions, which uses Fuzzy Reasoning Colored Petri Net (FRCPN), to verify the Web service compositions.

In [11], the author calculated the response time of clients, this performance was monitored using a monitoring Tool (pingdom).

in [12], the authors show how simple existing web services can be composed, in order to create a composite service, which offers new features. In this context, they propose an expressive object-oriented Petri Nets based algebra that succeeds in the complex composition of Web services.

in [13], the authors considered the possible compositions of Web services, and find all the services that optimize some attributed QoS under given constraints QoS that are NP-hard. As in reality, there are many Web services running at the same time, therefore, the authors modified the problem with consideration a repeatition of the same Web service. The authors modeled the problem modified as a linear program, and calculated the number of Web services task. In [14], The authors have developed a model of Web services using Unified Modeling Language (UML), Use Case Diagram, Sequence Diagram, Deployment Diagram. They obtain the Performance metrics by simulating the web services model using a simulation tool of Multi-Tier Queuing Architecture (SMTQA), they calculated: average response time, average waiting time, average Service time, probability of idle server, probability of dropping of sessions.

III. Problem setting

The Quality of Service (QoS) properties of service components and compositions are critical for their usage. Service Level Agreements (SLAs) are a means for defining permissible values for QoS attributes that are relevant in some scenario or for a particular purpose (such as execution time, monetary cost, or availability) and that a service (composition) provider is expected to deliver to a client.
The goal of this paper is to evaluate a system of clients requests satisfaction taking into account explicitly:

- The architecture characteristics, considering in term of performance evaluation;
- The traffic model, describing the requests traffic characteristics;
- Several causes of requests loss due to the dissatisfaction of a request by a simple and composite Web service.

Analytical performability models and simulation are developed to analyze the impact of the above aspects on the request's satisfaction.
The evaluation of the satisfaction of the requests is carried out taking account of the architecture of the system and the characteristics of the trafic. The proposed model describes the Web services system with Petri nets, where the arrival of requests from clients and Web services are taken into account to evaluate these performances.
A Petri net is a directed bipartite graph, composed of two types of nodes: places and transitions. Graphically places are represented by circles and transitions by lines or rectangles. Places and transitions are connected by arcs. Each places contains an integer number ($n \geq 0$) of tokens [15]. The state of the system modeled by a RdP, is represented by the marking of the network which is a vector, and which gives the distribution of the tokens in the places of the network [16].

IV. Web services discovery and composition: definitions and architecture

As commonly assumed in the literature, a Web service is a component of distributed applications which provide data to other applications by using the communication infrastructure offered by the Web.

It respects some properties such as autonomous object component, slightly coupled, self-describing, synchronous and asynchronous. To understand the Web services composition, we combine several Web services to create a composite Web service. According to Gardarin [17], Web services composition is a technique which assembles Web services in order to achieve a particular goal, via primitives of control (Test,Treatment of Exception,...) and Exchange (Sending and Reception of Messages).
According to Fensel [18], composition is a process which functions in an intelligent way in order to discover services automatically, and allows them to be combined in a more complex way.
We can say that Web services composition is the use of relevant Web services offered by various providers in order to obtain a composite service able to satisfy a user's request which cannot be satisfied by a simple available Web service.
An automatic and dynamic Web services composition is a highly complex task, because the proposed standards (XML,WSDL,UDDI,SOAP) of Web services technology do not answer the problems of Web services discovery and composition by a software agent. Also, the semantic annotations of Web services and requests are not yet mature.

1) **XML** [19]
   XML is a standard of W3C is a universal model of data representation and exchange.It is a simple format text, flexible and also independent of any manufacturer.Adding to this,it gives structure to documents and data(Harold and Means,2004). It is extracted from SGML language and it benefits from experiences of HTML's use(Hyper Text Mark-Up Language).Further more,XML offers portable and structured data on heterogeneous structure and programming languages.XML brings the following criteria to XML Web services architecture:

   - Extensibility: A system can function correctly without losing its main properties during an update.
   - Neutrality: The required constraints of an application are limited.
   - Structure: XML represents both document structure and content, offers increased control of information granularity through transformation and query languages.
   - Interoperability: Communication and data exchange between heterogeneous systems are possible.

2) **SOAP: simple object access protocol** [19]
   The SOAP protocol is an exchange message's process in heterogeneous environments for application-to-application communication based on XML and on standard protocol HTTP(Kadima and Monfort, 2004). SOAP,a standard of W3C,defines a set of rules to structure dialogs RPC(Remote Procedure Call)to exchange data. It ensures interoperability between components independent of transport mechanisms,operating systems and programming languages. According to Harold and Means(2004), SOAP is a flexible protocol to connect distributed systems.The purpose of this protocol is to facilitate the access to software services to any user through the Internet.

3) **WSDL: Web services description language** [19]
   WSDL is a formal language of Web services description according to the standard XML. A WSDL file describes the functionality (Methods, Parameters)and the localization of a Web service(URI, Port, Protocol of invocation).According to W3C(http://www.w3. org/TR/wsdl), WSDL separates the description of abstract functionalities offered by a service from concrete details of service description.As in programming languages,a type signature defines the inputs and outputs for a function.It means that WSDL can be seen as a traditional function,subroutine or method.
4) **UDDI: universal description, discovery and integration** [19]
   The registry of Web services "UDDI" is a virtual data base of existing XML Web services.It is similar to a CORBA trader and can be considered as a DNS service for business applications.On one hand, it allows providers of services to record XML Web services under a standardized format and on the other hand,it concentrates on discovery process of XML Web services satisfying services' needs in SOA. UDDI becomes an intermediate standard between providers and clients through the Internet and it is a recommendation of W3C.According to Chappell and Jewell (2002), the UDDI project is an initiative of industry which tries to create an independent platform, to describe services,to discover businesses and to integrate services.It means that UDDI provides a universal registry for business to provide service listings(Web service description).

V. Modeling requests satisfaction in Virtual Web services

The evaluation of the requests satisfaction is carried out taking into account the architecture of the system and traffic characteristics. We propose a model with

two stations: The first represents a client request and the second a Web services request, with one exponential server. Arrivals to the network follows the Poisson distribution, and the arrival rates are λ_i, $i \in 1, 2$ and are independents. When the clients request and Web services come to the system, they are placed in the place $P_1^{'}$, $P_2^{'}$ respectively. We select n clients to come to the place P_1 and m Web services to come to the place P_2, after crossing the transitions T_1 and T_2 respectively. Then, they cross the transition T with a rate μ to serve the client request. When the service (discovery and composition) is completed, the Web services requests rejoin their station. Service rate (discovery and composition) is μ. Fig. 2 presents a Petri nets model of a Web services system. The meanings of places

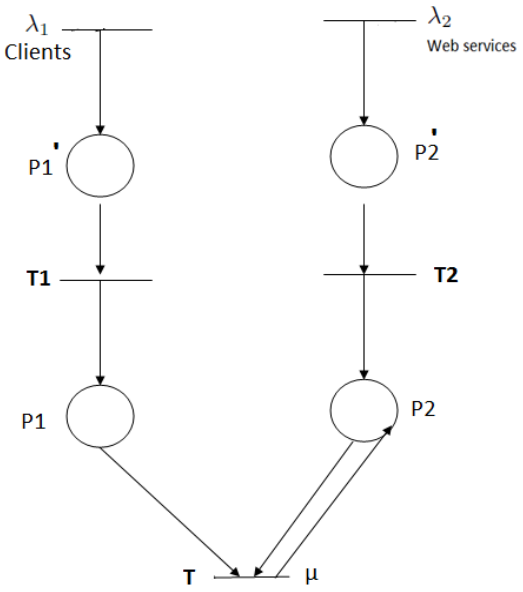


Figure 2. Petri Nets model

$(P_1^{'}, P_2^{'}, P_1, P_2)$ and transitions (T_1, T_2, T) in the model are:

- $P_1^{'}$: Clients queue, unlimited place.
- $P_2^{'}$: Web services queue, unlimited place.
- P_1: Clients queue, place with limited capacity.
- P_2: Web services queue, place with limited capacity.
- T_1: Selection of a clients requests.
- T_2: Selection of a Web services.
- T: Clients service.

VI. Model analysis

We proposed a Petri nets model of a Web services system, see Fig. 2. In using simulation by the GRIF simulator, the performance indices were calculated.

In order to calculate average response time and average number of clients in the system, we assume that the inputs criteria are: $\lambda_2 = 0.01$, $\mu = 1$, the different values of λ_1 are:
$\lambda_1 = 0.00001, 0.0001, 0.001, 0.01, 0.1$ and the different values of n are:
$n = 1, 10, 100, 10000, 100000, 1000000, 10000000$.

VII. Simulation results

After implementing the model in the GRIF simulator, we were able to simulate our system according to the differents values, we recovered the simulation results.
Tab. I presents the average response time as a function of n, and differents values of arrival rate λ_1.
From Tab. I, we note that the average response time

Table I. Average response time.

	$\lambda_1 = 0.00001$	$\lambda_1 = 0.0001$	$\lambda_1 = 0.001$	$\lambda_1 = 0.01$	$\lambda_1 = 0.1$
$n = 1$	19572.83	21900.07	25449.28	27714.07	29394.33
$n = 10$	19533.75	25744.56	21789.18	27785.13	29544.21
$n = 10^2$	19681.14	25668.71	21952.35	28078.93	30254.03
$n = 10^3$	19779.39	26089.77	22026.26	28637.30	38502.57
$n = 10^4$	29471.32	31758.26	35600.44	37634.52	38502.57
$n = 10^5$	29586.24	31871.57	35706.82	37668.84	38502.57

increases for each increase of λ_1, because the number of arrivals increases. Also each time n increases, the average response time increases.
Tab. II presents the average number of clients in the system as a function of n, and differents values of arrival rate λ_1.

Table II. Average number of clients in the system

	$\lambda_1 = 0.00001$	$\lambda_1 = 0.0001$	$\lambda_1 = 0.001$	$\lambda_1 = 0.01$	$\lambda_1 = 0.1$
$n = 1$	49.13	49.75	51.20	57.11	67.92
$n = 10$	57.59	58.22	59.91	66.38	78.11
$n = 10^2$	150.73	151.23	153.02	158.14	166.99
$n = 10^3$	1092.91	1093.42	1095.87	1102.12	1120.55
$n = 10^4$	14989.94	14991.65	14996.50	15036.88	15496.88
$n = 10^5$	194989.17	194990.90	194995.82	195036.77	195496.88

From Tab. II, we note that the Average number of clients in the system increases for each increase of λ_1, because the number of arrivals increases, because we give clients more of a chance to submit their requests. Also each time n increases, the average number of clients in the system increases.

VIII. CONCLUSIONS

In this work, we modeled a system of Web services with Petri nets, which we considered the arrivals of Web services and clients in the system. We calculated the average number of clients in the system and the average response time, because we give clients more of a chance to submit their requests. Hence we will have the average response time which increases. According to the results found, we remark that a some time, the system saturates and cannot receive more clients. as a perspective, we propose to do an analysis, in the case where the Web services system is saturated.

REFERENCES

[1] H. Kadima, "*Web services: techniques, procedures and tools XML, WSDL, SOAP, UDDI*," RosettaNet, UML. Dunod, 2003.

[2] H. Talantikite, "*Complex Web services: Discovery, composition, selection, and orchestration*," Doctoral Theses in Computer Science, Abderahmane Mira of Bejaia university, Algérie, 2010.

[3] N. Bernine et al, "*Towards a performance analysis of composite Web services using Petri nets*," International Journal of Mathematics in Operational Research, Vol. 17, No. 4, p.467-491 (2020).

[4] Haddad, S et al, "*Bounding models families for performance evaluation in composite Web services*," Computational Science Journal, Vol. 4, No. 4, pp.232–241 (2013).

[5] Garima, G et al, "*Performance Evaluation of Composed Web Services*," International Journal of Computer Applications, Vol. 171, No. 4, pp.975–8887 (2017).

[6] Chattopadhyay, S and Banerjee, A, "*QoS constrained Large Scale Web Service Composition using Abstraction Refinement*,"IEEE Transactions on Services Computing, No. 99, pp.1-1 (2017).

[7] Singh, R.P and Pattanaik, K.K, "*An Approach to Composite QoS Parameter based Web Service Selection*," Proceedings of the 4th International Conference on Ambient Systems, Networks and Technologies (ANT 2013), Halifax, Nova Scotia, Canada, pp.470–477 (2013).

[8] Mokdad et al, "*Bounding models families for performance evaluation in composite Web services*," Proceedings of the 1st International Conference on Industrial Networks and Intelligent Systems (INIS'15), Tokyo, Japan (2015).

[9] Rosenber, F et al, "*Metaheuristic Optimization of Large-Scale QoS-Aware Service Compositions*," Proceedings of the IEEE International Conference on Services Computing, Miami, Florida, pp.97-104 (2010).

[10] Deng, Z et al, "*A Reliability Calculation Method for Web Service Composition Using Fuzzy Reasoning Colored Petri Nets and Its Application on Supercomputing Cloud Platform*," Future Internet Journal, Vol. 8, No. 4, pp.47-47 (2016).

[11] Kumar, M, "*Various Factors Affecting Performance of Web Services*," International Journal of Sensor and Its Applications for Control Systems, Vol. 3, No. 2, pp.11-20 (2015).

[12] Chemaa, S et al, "*A high-level Petri net based approach for modeling and composition of web services*," Proceedings of the International Conference on Computational Science (ICCS), Omaha, Nebraska, pp.469 - 478 (2012).

[13] Klein, A et al, "*Efficient QoS-aware Service Composition with a Probabilistic Service Selection Policy*," Service-Oriented Computing Journal, Vol. 6470, pp.182-196 (2010).

[14] Reddy, Ch R. M et al, "*Early Performance Prediction of Web Services*," International Journal on Web Service Computing (IJWSC), Vol. 2, No. 3, pp.31-41 (2011).

[15] C. A. Petri, "*Interpretations of net theory*," GMD,Technical Report, 1976.

[16] R. David et Al, "*From Grafcet to Petri nets*," Hermes, 1992.

[17] G. Gardarin, "*XML des bases de données aux Services Web*," Dunod, 2002.

[18] D. Fensel et Al, "*Semantic Web Enabled Web Services*," Sigmod Record ACM, Vol. 31, No. 4, pp.1-2 (2002).

[19] Nacer, H and Aïssani, D, "*Semantic web services: Standards, applications, challenges and solutions*," Network and Computer Applications Journal, Vol. 44, pp.134-151 (2014).

Towards a reliable multi user full duplex protocol in high efficiency WLANs

Hocini Kenza
Research Unit LaMOS, Faculty of Exact Sciences, University of Bejaia, Algeria
kenza.hocini@univ-bejaia.dz

Yazid Mohand
Research Unit LaMOS, Faculty of Exact Sciences, University of Bejaia, Algeria
mohand.yazid@univ-bejaia.dz

Abstract—
To enhance the network capacity and spectral efficiency in HEW (High Efficiency WLANs). The standard IEEE 802.11ax introduced the Full Duplex transmission technique, which enables at the station to transmit UL (Up-Link) data and receive DL (Down-Link) data from the access point at the same time and on the same radio frequencies. However, two challenges remain to be met in order to propose a reliable MAC protocol which are: (1) the asymmetry between the amounts of DL and UL frames. And (2) detection of hidden station in order to handling the opportunity window. The main goal of this paper is to design and propose the a reliable MU-UL-FDRC (Multi User Up Link Full Duplex Radio Communications) protocol by meeting each of challenges. The performance evaluation demonstrates that MU-UL-FDRC protocol enhances the overall simulation results.

Keywords—
high efficiency WLANs, Full Duplex, Multi User, MAC protocol, Performance Evaluation.

I. Introduction

IEEE 802.11 is part of the IEEE 802 set of local area network (LAN) technical standards, and specifies the set of media access control (MAC) and physical layer (PHY) protocols for implementing wireless local area network (WLAN) computer communication. The standard and amendments provide the basis for wireless network products using the Wi-Fi brand and are the world's most widely used wireless computer networking standards [1]. IEEE 802.11 is used in most home and office networks to allow laptops, printers, smartphones, and other devices to communicate with each other and access the Internet without connecting wires. Actually, the IEEE has started a new working group called 802.11ax with the focus on multi-cell deployment in a dense environment, such as stadiums, train, apartment building, shopping malls, conference venues, and other outdoor scenarios [2]. IEEE 802.11ax was approved in March 2014 to design a brand new amendment for the new generation WLANs [3]. IEEE 802.11ax requires four times throughput improvement in dense deployment scenarios compared with the current IEEE 802.11 WLAN and the nominal data rates are increased up to 9.6 Gbps [4]. However, simply increasing the bandwidth for higher throughput will no longer be available since the spectrum resource is increasingly scarce. Therefore, new wireless communication technologies are needed to improve the spectrum efficiency in the next generation WLAN [5].

The Full Duplex (FD) radio technology has been spotlighted as one of the key technologies in the future generation of High Efficiency WLANs (Wireless Local Area Networks). Indeed, it has the advantage of doubling the throughput of the Half Duplex radio (HD) through the adoption of SIC (Self Interference Cancellation) antennas. By means of said antennas, simultaneous transmissions and receptions are therefore possible at the same time on the same radio frequencies [6].

However, two challenges remain to be met in the full duplex radio communications as follows: (1) the asymmetry between the amounts of DL and UL frames. And (2) detection of hidden station in order to handling the opportunity window.

In this paper, we propose a reliable multi user full duplex protocol in IEEE 802.11ax. It based mainly on multi user access by utilizing the full duplex transmission technique defined in the IEEE 802.11ax standard, which enables to multiple data streams UL and aggregated frame DL to be simultaneously transmitted by the multiple hidden stations to the AP, and from the AP to hidden stations respectively, on the same radio frequencies without interference in order to efficiently exploit the opportunity window caused by the asymmetry between the amounts of DL and UL frames. Furthermore, the obtained numerical results illustrate the robustness of the proposed protocol in terms of UL throughput, service rate and overhead metrics.

The remainder of this paper is organized as follows. In Section II, we will describe the full duplex communications, and we will highlight challenges inherent to their operation. In Section III, we will outline the important full duplex MAC protocol in high efficiency WLANs available in the literature. In Section IV, we will describe our proposal for multi user. In Section V, we will present and analyze the simulation results of the proposed protocol. In Section VI, we will conclude the paper, and we will suggest some research perspectives.

II. Background

In this section, we describe the asymmetry issue in the Sub-section II-A, and the behavior of stations in the Sub-section II-B.

A. Asymmetry issue

The fact that the access point is the relay station in the network. Thus, the DL data streams descending from the access point is most voluminous data stream compared to the UL data stream ascending from station towards the access point. This imbalance known as asymmetry issue [7]. Due to this imbalance between the lengths of DL and UL frames, an Opportunity Window (OW) is generated as illustrated in Figure 1. It represents the difference between the transmission time of the DL frame and that of the UL frame. This empty time interval impedes an effective use of full duplex radio technique. For this; the hidden stations can use this opportunity window. We will describe the problem of hidden stations in the following sub-section.

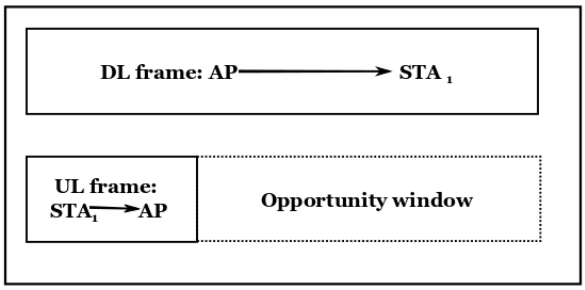


Figure 1. The asymetry between DL and UL frames.

B. Behavior of stations

The Figure 2 illustrates the problem of the hidden stations. We have previously the Bidirectional Full duplex (BFD) between the access point (AP) and a station (STA); thus, the downlink traffic from the AP towards a STA1 and uplink traffic from the STA1 towards the AP. We call STA2 is hidden to STA1, because it receives correctly the downlink transmission destined to STA1, at the same as STA1 sends uplink transmission towards the AP. But, STA3 is exposed to STA1, it cannot hear the downlink traffic from the access point towards STA1, because of interference which is caused by uplink traffic. Even with full duplex, STA3 cannot receive two frames simultaneously.

We summarize that to find stations with UL traffic in the hidden relation with the station for downlink traffic, it would have previously bidirectional full duplex transmission between an access point and a station and If a station in the network can receive DL transmission without interference to the current UL transmission, then it is a hidden station. However, if a station can not receive DL transmission, then it is an exposed station [8].

To fill the imbalance length between downlink and uplink traffics, we should find the hidden relation among stations.

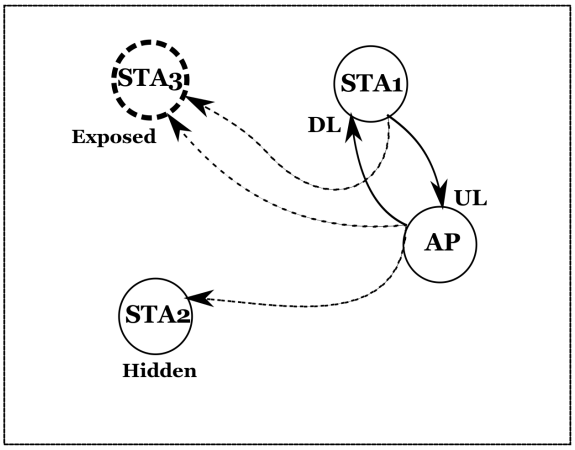


Figure 2. Behavior of stations.

III. Related work

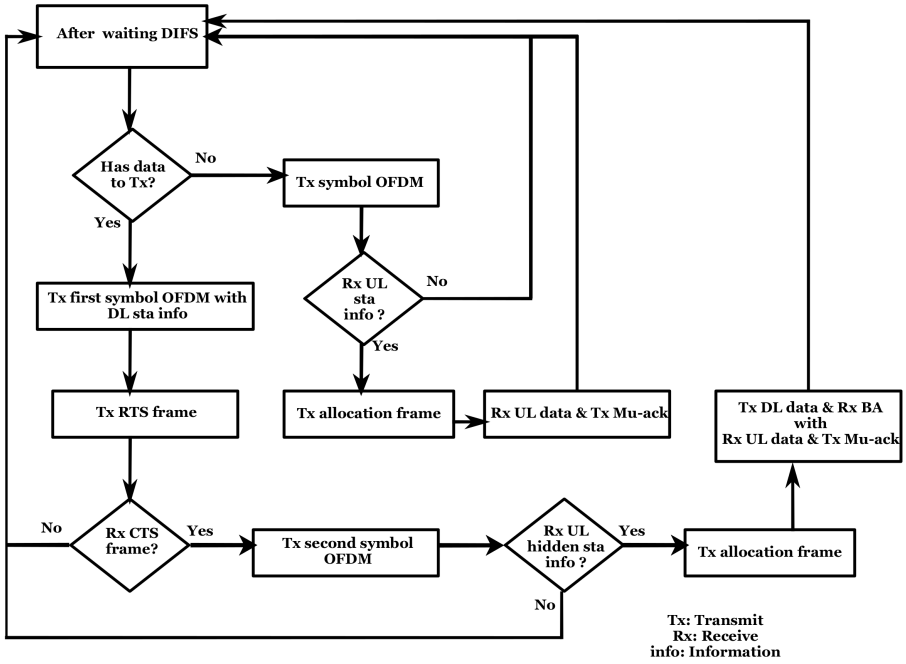
In this section, we describe the most existing studies dealing with full duplex radio communications.

Alim et al. in [9] proposed an IBFD MAC protocol named Asym FDMAC for infrastructure based WLAN to support asymmetric lengths of traffic for uplink and downlink. This MAC protocol enables multiple users to transmit data to the access point (AP) during the transmission of a single downlink frame from the AP. In Asym-FDMAC, the AP always initiates the transmission. Therefore, there is no contention period and thus, there is no collision among the user terminals and the AP for channel access.

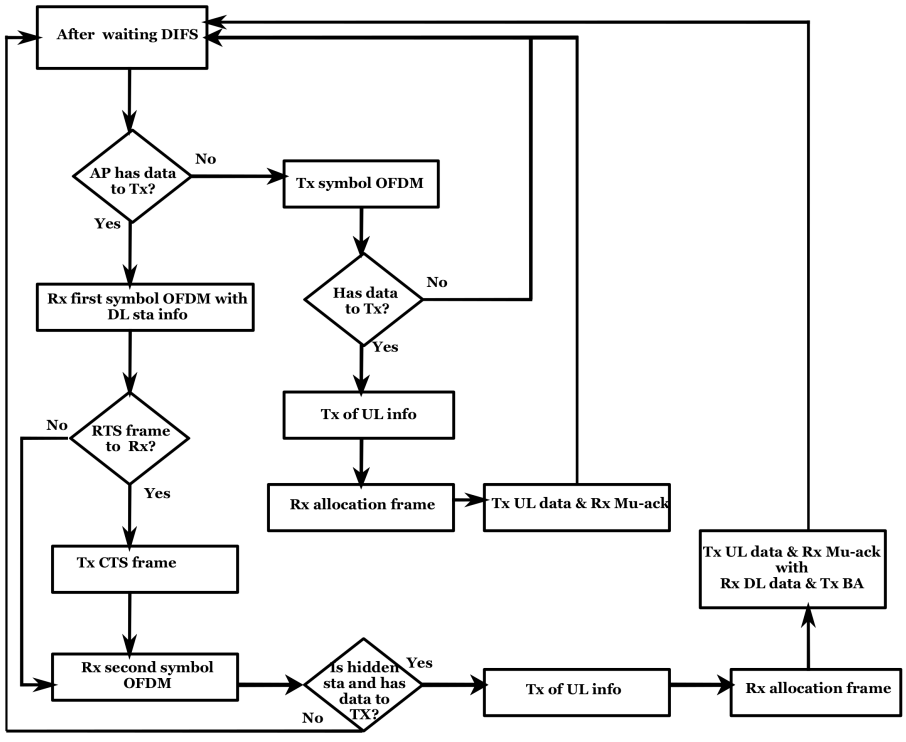
Alkhrijah et al. in [10] proposed a new protocol FD-MUMAC (Medium Access Control (MAC) protocol for Full Duplex Multi User multi input multi output), which determines the selection of uplink and downlink users and controls the transmission rate by considering transmit and receive beamforming, channel states, and multiuser interferences.

Hyeongwoo et al. in [11] proposed a novel medium access control scheme, called FDCR, which resolves a collision between full duplex and legacy stations. FDCR provides collision resolution without modifying the 802.11 standard of legacy stations. When a collision occurs, the full duplex station reports information on the colliding stations to the access point. Then transmission opportunity is granted to them.

Hocini et al. in [12] proposed ECC FDRC protocol (Efficient and Coordinated Control of Full Duplex Radio Communications) by tacking into account four



(a) For AP.



(b) For station.

Figure 3. State transition diagrams of MU-UL-FDRC protocol.

challenges, namely: Full-Duplex configurations, hidden relationship, unbalance between data streams, and overhead. The basic principle is to exchange all relevant control information and transmit data without loss or unnecessary delay.

Yazeed et al. in [13] proposed a HyFDMAC (Hybrid Full duplex Medium Access Control) protocol that integrates random and scheduled access for infrastructure based wireless networks. HyFDMAC consists of multiple stages based on IEEE 802.11 distributed coordination function and enables a full duplex access point to collect the channel and interference information required for multiuser transmission. HyFDMAC guarantees fairness using the random and scheduled access mechanisms.

IV. Proposed solution

In this section, we propose a multiple access method for UL stations based on Full Duplex technology in high efficiency WLANs, which we call reliable MU-UL-FDRC (Multi User Up Link Full Duplex Radio Communications) protocol. It enables to optimize the number of UL stations in the network with high connection demand (massive network connection), in order to increase the UL throughput and decrease the overhead in the network. One communication cycle of MU-UL-FDRC protocol consists of two phases, namely: signalization phase and transmission phase.

The MU-UL-FDRC protocol operates in a 20 MHz transmission channel, and we assume that a network which consists of a single AP and multiple stations associated with the AP employs a Full Duplex radio. The MU-UL-FDRC protocol works according to the state diagrams as shown in Figure 3. State transitions for AP and stations are illustrated in Figure 3(a) and Figure 3(b), respectively.

When the channel is idle, the AP waits DIFS (Distributed Inter Frame Space), if it has data to send, it transmits the first OFDM symbol in order to inform the designated station. Then, the exchange of RTS and CTS frames between the AP and the designated station in bidirectional full duplex mode for the reason of determine whether the station in the network is hidden or exposed to the designated station. After that, the AP sends the second OFDM symbol in order to invite hidden stations to send their needs in terms of UL information. According to the UL information, the AP sends allocation frame and then sends DL data, and multi user acknowledgement (Mu-Ack) and receives UL data and bloc acknowledgement (BA). If the AP has no data, it also transmits OFDM symbol in order to invite all stations in the network to send their needs in terms of UL information, then it sends allocation frame. Finally, the AP receives UL data and transmits Mu-Ack.

V. Performance evaluation

This section is dedicated to implementing, simulating and evaluating the performance of the protocol proposed in terms of three metrics of performance which are: throughput of the amount UL, the service rate, and overhead in function of the number of stations under Matlab. The following simulation results have been obtained by assuming the used parameters listed in the Table I.

Table I. Simulation Parameters

Parameter	Numerical Value
Channel bandwidth	20 Mhz
Basic rate	8.6 Mbps
Data rate	143.4 Mbps
DIFS	34 μs
PHY header	40 μs
MAC header	288 bits
RTS frame	160 bits
CTS frame	112 bits
OFDM symbol	12.8 μs
Guard inetrval	0.8 μs
MPDU length	11454 bytes

In Figure 4, we evaluate the variation of the UL throughput according to the number of stations in the network of the MU-UL-FDRC protocol. We observe that the UL throughput increases proportionally to the size of the network. Indeed, more the number of stations in the network increase, more parallel transmissions of several stations at the same time, more the amount of UL data is higher, which allows the UL throughput to increase.

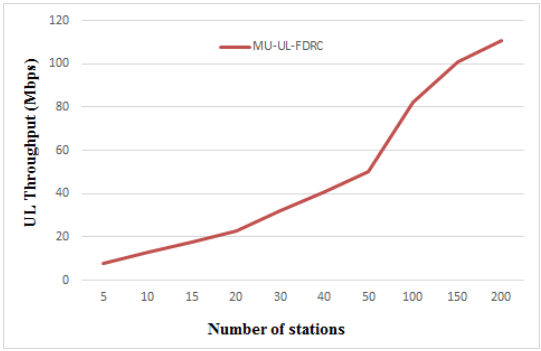


Figure 4. UL throughput in function the number of stations.

In Figure 5, we evaluate the variation of the UL throughput according to the number of stations in the network of the MU-UL-FDRC protocol. We see that, the service rate increases with increasing network size, this comes down to the fact that the proposed protocol uses the multi-user technique, where several stations can access the channel in parallel transmission, which increases the number of users served for a single transmission.

In Figure 6, we demonstrate the variation of the overhead induced mainly by asymmetry between the DL and UL data streams in function of the number of stations in the network. We see that, when the number of stations

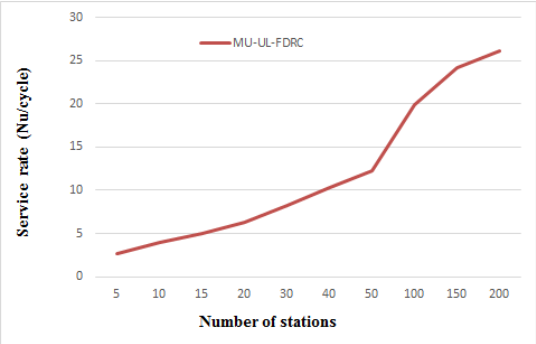


Figure 5. Service rate in function the number of stations.

increases, the overhead decreases, this comes down to the fact that the padding time in the MU-UL-FDRC protocol is much less, this is due to the increase of the amount UL data stream. In other words, the difference between DL and UL data streams is reduced with the increase in the number of stations.

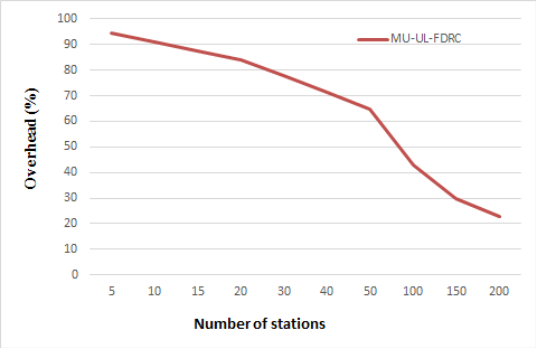


Figure 6. Overhead in function the number of stations.

VI. Conclusion and Prospects

In this paper, we focused on multi user access by utilizing the full duplex transmission technique defined in the IEEE 802.11ax standard, which enables to multiple data streams UL and aggregated frame DL to be simultaneously transmitted by the multiple hidden stations to the AP, and from the AP to hidden stations respectively, on the same radio frequencies without interference in order to efficiently exploit the opportunity window caused by the asymmetry between the amounts of DL and UL frames. We have proposed a reliable multi user full duplex protocol in HEW networks, in order to solve the two issues: (1) the asymmetry between the amounts of DL and UL frames. And (2) detection of hidden station to exploit this latter. We implemented and simulated the protocol proposed in terms of throughput UL, service rate, overhead in function of the number of stations. In the continuity of this research work, we would like to compare the protocol proposed with other available in literature.

References

[1] S. Mammeri, M. Yazid, L. Bouallouche-Medjkoune, A. Mazouz, "Performance study and enhancement of multichannel access methods in the future generation VHT WLAN," *Future Generation Computer Systems*, 79, pp 543-557, 2018.

[2] IEEE. Draft standard for information technology-telecommunications and information exchange between systems local and metropolitan area networks - specific requirements part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment enhancements for high efficiency WLAN. P802.11ax/D6.0, IEEE Standards Activities Department, 2019.

[3] S. Brahmi, M. Yazid, "Towards a Fair Allocation and Effective Utilization of Resource Units in Multi-User WLANs-based OFDMA technology," *Computer Networks*, 224, pp 109639, 2023.

[4] K. Hocini, M. Yazid, "Full duplex radio communications in high efficiency WLANs: Study and comparison of the main MAC protocols," *International Journal of Informatics and Applied Mathematics*, 3(1), pp 1-21, 2020.

[5] K. Hocini, M. Yazid, "Towards high performance full duplex MAC protocol in high efficiency WLANs," *In Artificial Intelligence and Renewables Towards an Energy Transition*, pp 756-765, 2021.

[6] D. Bharadia, E. McMilin, S. Katti, "Full duplex radios," *In Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, pp. 375-386, 2013.

[7] K. Hocini, M. Yazid, "Towards a Full-Duplex MAC protocol efficiently handling the imbalance between the lengths of data streams in High-Efficiency WLANs," *In 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)*, pp 148-153, 2021.

[8] K. Hocini, M. Yazid, M. "Toward a MAC protocol overcoming hidden stations issue in IEEE 802.11 ax unidirectional full duplex radio communications," *Second International Conference on Embedded and Distributed Systems (EDiS)*, pp 111-116, 2020.

[9] M. Alim, S. Saruwatari,T. Watanabe, "Asym-FDMAC: In-band full-duplex medium access control protocol for asymmetric traffic in wireless LAN," *Wireless Networks*, 26, pp 807-822, 2020.

[10] Y. Alkhrijah, J. Camp, D. Rajan, "Full duplex multiuser mimo mac protocol (fd-mumac)," *IEEE Global Communications Conference*, pp 1-6, 2020.

[11] H. Jo, H. Ahn, E. Kim, Y. J. Suh, "FDCR: A Full-Duplex Collision Resolution Scheme for Next-Generation Wireless LANs," *IEEE Communications Letters*, 25(11), pp 3738-3742, 2021.

[12] K. Hocini, M. Yazid, A. Ksentini, "Toward an efficient and coordinated control of full-duplex radio communications in high-efficiency wireless local area networks," *International Journal of Communication Systems*, 35(10), pp e5165, 2022.

[13] Y. Alkhrijah, J. Camp, D. Rajan, "HyFDMAC: A Hybrid Access Full-Duplex MAC Protocol," *15th International Conference on Communication Systems and NetworkS (COMSNETS)*, pp 572-578, 2023.

Éléments de Supervision, de fiabilité et d'optimisation des réseaux mobiles à Béjaia

Tounsi Mohamed
Lamos Research Unit, UAMBéjaia
mohamed.tounsi@univ-bejaia.dz

Ouazziz Yacine
Limed Laboratory, UAMBéjaia
yacine.ouazziz@univ-bejaia.dz

***Résumé*—**
La supervision d'un réseau radio mobile est une tâche complexe vu qu'elle doit intervenir à différents niveaux : la surveillance continue et l'optimisation des performances, la réalisation répétée de statistiques, la détection et la correction d'anomalies ou pannes de fonctionnement, la reconfiguration et le paramétrage des équipements, ainsi que la maitrise de leur fiabilité par des actions adéquates de maintenances préventive et corrective. Cet article est une revue de plusieurs investigations sur l'optimisation des performances de diverses technologies de réseaux mobiles déployés à Béjaia, ainsi qu'une analyse statistique sur la fiabilité et la disponibilité de sites radiomobiles.

Mots-Clés—
Indicateurs KPI, Optimisation de performances, Qualité de service (QoS), Gestion d'alarmes, Détection de pannes

I. INTRODUCTION

Les communications sans fil ont connu une vraie révolution depuis 1990, avec l'apparition de divers standards (WiFi, Bluetooth, WiMax...) et le déploiement rapide de générations successives de réseaux cellulaires (2G, 3G, 4G). Diverses avancées technologiques ont ainsi permis de passer de réseaux analogiques à des réseaux numériques quadruple play (voix, données, vidéo, mobilité) avec une connectivité à des débits élevés, des temps de latence réduits, une QoS et une fiabilité accrues. Des développements récents (5G, 6G) visent une transition numérique de l'industrie et des services [1, 2].

Les réseaux cellulaires font face à une multitude de contraintes, notamment liées à leur interface radio et à la limitation des ressources spectrales à partager sur les utilisateurs. Cela exige des opérateurs, une gestion supervisée en vue d'optimiser leurs performances et leur rentabilité. Dans ce qui suit, nous synthétisons plusieurs travaux menés sur divers aspects de la supervision des systèmes mobiles déployés dans la région de Béjaia.

II. ARCHITECTURE GENERIQUE DES RESEAUX MOBILES

L'architecture des réseaux mobiles a été construite de façon modulaire et évolutive, avec une organisation hiérarchique et diverses fonctionnalités héritées du réseau public RTCP. Malgré une tendance actuelle vers de nouvelles topologies sur un réseau tout IP avec des équipements génériques implémentant des fonctions logicielles de commutation et de gestion de base de données et même des techniques de virtualisation, d'élasticité et de slicing, la structure de base d'un réseau mobile est illustrée par la Fig.1 [3, 4].

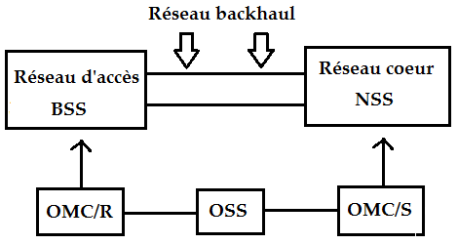


Figure. 1. Structure génériques des réseaux mobiles.

Calquée sur l'architecture de référence (GSM), cette structure est constituée d'un réseau d'accès BSS *(Base Station SubSystem)* intégrant une technologie radio, d'un réseau cœur NSS *(Network Switching SubSystem)* assurant les fonctions essentielles du réseau, reliés par un réseau d'amenée *(backhaul)* pour joindre les différentes entités et collecter le trafic usager et la signalisation. La gestion et la supervision du réseau sont assurées par le centre OSS *(Operations support SubSystem)* et le centre NMC *(Network and Management Center)* subdivisé en centre OMC/R *(Operations and Maintenance Center/Radio)* relié à toutes les entités du BSS, à travers les BSC, et l'OMC/S (System) qui est relié au NSS à travers les MSC comme montré en Fig.2 [4].

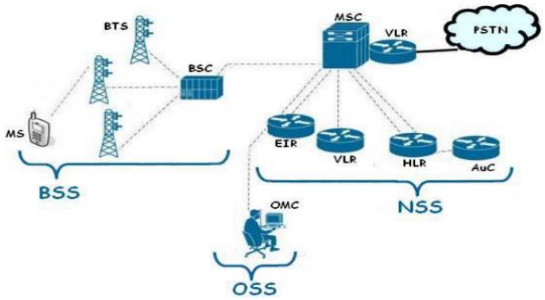


Figure. 2. Architecture de référence du réseau GSM.

III. METHODOLOGIE DE SUPERVISION ET D'OPTIMISATION DES RESEAUX MOBILES

A. *Organisation fonctionnelle*

L'implémentation de la supervision est laissée avec beaucoup de liberté par la norme GSM mais classiquement, l'organisation fonctionnelle de cette tâche est structurée comme illustré par la Fig.3.

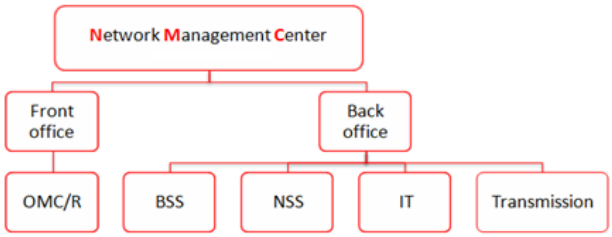


Figure 3. Organisation fonctionnelle des taches de supervision [5].

La supervision se fait à travers des canaux dédiés, par plusieurs équipes d'ingénieurs complémentaires qui sont chargés de surveiller les performances du réseau (coté BSS et coté NSS), de récolter et régler des anomalies (alarme, Drive-tests, réclamations, Benchmarking ...) par des actions correctives à distance ou par des ingénieurs de maintenance sur terrain.

B. *Niveaux de supervision*

La supervision de l'état de fonctionnement d'un réseau mobile intervient à différents niveaux : la récolte continuelle de divers indicateurs clé KPIs de performances, la récolte d'alarmes selon plusieurs degrés de sévérité, la réalisation d'enquêtes et de drive-tests suite à des plaintes, réclamation ou constatation de baisse de performance récurrente.

1) *Indicateurs de performances KPIs*

Les KPIs renseignent sur la QoS offerte aux abonnés. Plusieurs compteurs OMC sont disséminés dans l'infrastructure du réseau pour évaluer ces indicateurs qui différent selon les équipementiers et la technologie déployée. À titre d'exemple, la firme Ericsson définit pour ces équipements 4G/LTE des KPIs concernant ***l'accessibilité*** *(capacité d'obtenir un service dans une période limitée)*, **l'intégrité** *(capacité de fournir un service demandé par l'utilisateur)*, la **mobilité** *(capacité de fournir un service à un utilisateur en mouvement)* et la **disponibilité** *(capacité de répondre au service demandé par l'utilisateur)*[6-8]. L'analyse des mesures KPIs permet de corriger les anomalies constatées. La figure 4 illustre le cas d'une mauvaise accessibilité due à un mauvais taux d'accès aléatoires, ou à un mauvais taux d'établissement de sessions.

2) *Indicateurs de Drive-tests*

Les Drive-test sont des campagnes de mesures effectuées par des ingénieurs de l'opérateur, munis d'un matériel spécifique (chaine de mesure, GPS), à travers des zones de couverture ciblées pour tester les performances du réseau, réellement perçues par les abonnés.

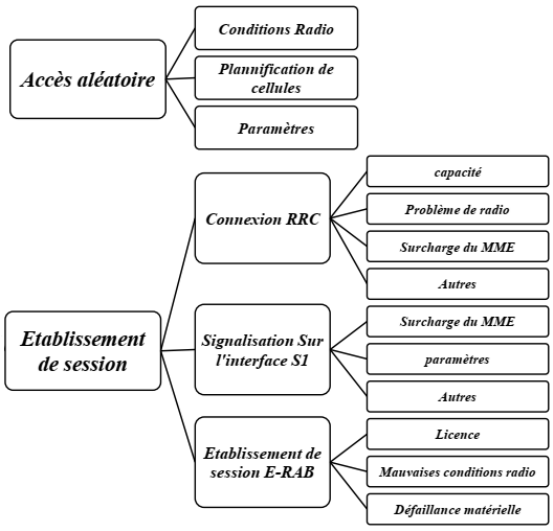


Figure 4. Causes de dégradation de l'accessibilité en 4G/LTE [5].

Les indicateurs utilisés lors des drive-test permettent de trouver des solutions d'optimisation, de densification ou de redimensionnement des sites où des baisses de performances sont vérifiées. Le tableau Tab.1 donne les indicateurs mesurés pour des sites 4G [9].

Tableau 1. Indicateurs de drive-test pour des systèmes 4G/LTE.

RSRP	puissance moyenne des signaux de référence reçus par l'UE en mode veille ou connecté
RSRQ	qualité du signal de référence reçu par l'UE en mode connecté
SINR	rapport signal sur bruit reçu par l'UE
RSSI	puissance totale reçue par l'UE, incluant même les interférences et le bruit
CQI	qualité du signal reçu par l'UE sur son canal
PCI	identifiant de la cellule actuelle (de 0 à 503)

3) *Processus d'optimisation après dégradation de KPIs*

Quand une baisse de performances est avérée suite à une dégradation d'indicateurs KPIs de QoS ou de Drive-tests, un processus d'optimisation est élaboré selon le cas pour corriger les performances du réseau. La figure 5 montre le cas d'une optimisation du réseau radio.

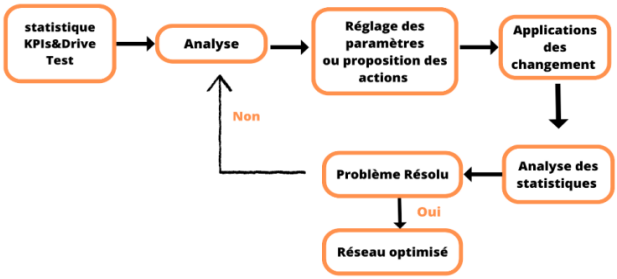


Figure 5. Étapes d'optimisation radio d'un réseau mobile.

4) Récolte d'alarmes

La gestion des anomalies d'un réseau mobile exige une surveillance constante de ces performances. Les centres OMC récoltent diverses alarmes qui seront filtrées pour être affichées au niveau du service concerné. Souvent, un automate s'occupe de la vérification automatique des sites périodiquement pour voir si les alarmes apparues ont été traitées, auquel cas, il ferme l'incident, sinon, des ordres de mission sont transmis aux équipes sur le terrain pour des actions correctives. Pour organiser le suivi, les opérateurs adoptent des plates-formes ou des outils logiciels leur permettant de recueillir, de documenter les alarmes survenues. L'opérateur Ooredoo utilise la plate-forme **HP/TeMIP** (*Telecommunications Management Information Platform*), tandis que l'opérateur Mobilis utilise un outil logiciel ***4T*** (**Technical Trouble Ticket Tool)** pour la gestion des incidents ainsi que les demandes de changements en temps réel.

IV. TRAVAUX REALISES

Dans cette section, nous présentons une revue de différentes études et investigations sur terrain menées sur plusieurs années sur des réseaux mobiles déployés à Béjaia.

A. *Évaluation de la QoS des réseaux mobiles*

1) Réseaux GSM Mobilis /année 2014-2015

Une première investigation a été menée en 1994 sur quelques sites GSM de la wilaya de Béjaia. On synthétise ici l'étude des KPIs d'accessibilité au réseau [6] selon les mesures relevées du taux de disponibilité et de congestion perçus par les usagers dans certaines cellules [10].

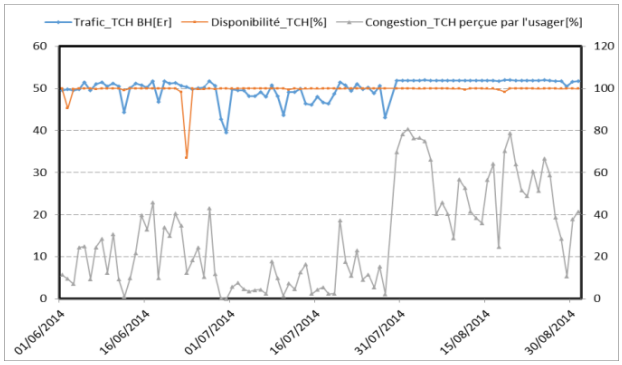


Figure 6. Disponibilité TCH, trafic TCH et congestion perçue par l'usager pendant l'été 2014 dans la cellule CB3 [10].

La figure ci-dessus montre pour l'une des cellules, une disponibilité de 100% pour un trafic avoisinant 50 Erlang, sauf pour la journée du 23/06/2014 où le manque de disponibilité atteint 66.7%, cette valeur est très inférieure au seuil. Après analyse, il s'est révélé que cette baisse est due à une panne et une réparation au niveau des TRXs. On observe aussi une forte congestion tout au long du trimestre alors que les canaux TCHs sont disponibles. La seule façon de remédier à cette congestion est d'augmenter la capacité du réseau.

La figure 7 représente le diagramme de trafic et de taux de disponibilité et de congestion TCH perçu par l'usager dans la macro cellule CB5 [10] .

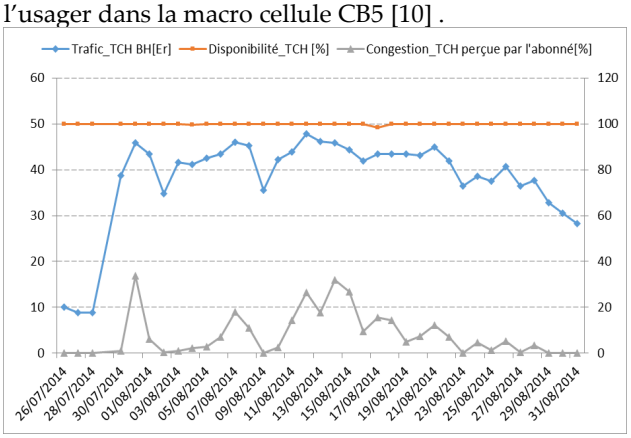


Figure 7. Diagramme de trafic et des taux de disponibilité et de congestion TCH dans la cellule CB5 [10].

On remarque pour cette cellule, que la disponibilité des ressources est de 100% tous le long du mois. Le trafic a augmenté subitement de 20 Er à 47Er, qui a entrainé une congestion TCH. Cela est dû à la période estivale, puisque la cellule est située sur le littoral, donc il y a une forte densité d'abonnés. Pour réduire cette congestion, une solution a été de placer temporairement une BTS mobile près de cette zone.

Sur la figure 8, nous donnons le taux de coupure sur une période d'une année. Nos mesures révèlent un taux inférieur au seuil sauf pour la journée du 12/06 ou il atteint 4.5% du probablement à une valeur excessive du timing advance ou à un problème de transmission.

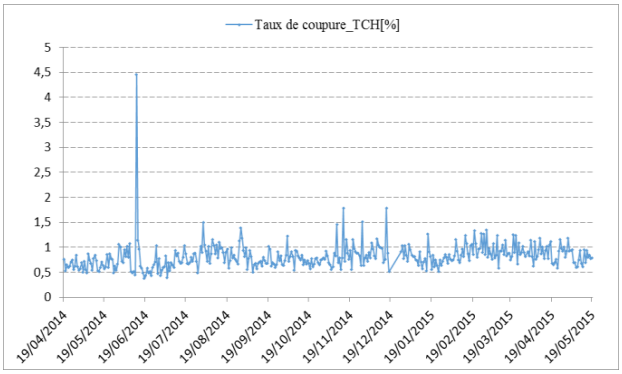


Figure 8. Taux de coupures TCH sur une année dans la cellule CB2 [10].

2) Réseaux UMTS Ooredoo /année 2016

Une autre étude a été menée sur la QoS du réseau 3G de l'opérateur Ooredoo durant l'année 2016. La figure 9 représente le diagramme du protocole de signalisation [7] pour une période de cinq mois dans la région de Bejaia.

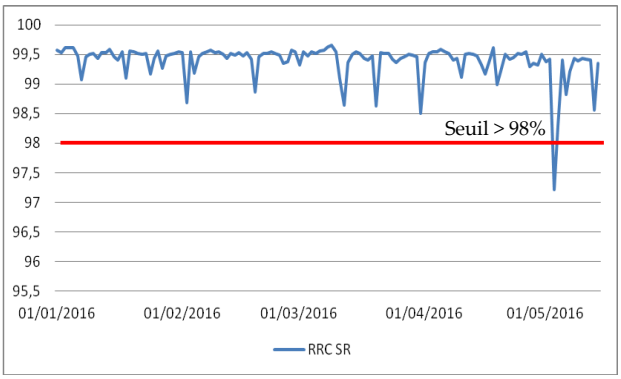


Figure 9. Taux de connexion RRC sur 5 mois dans une cellule [11].

D'après cette figure, le taux RRC SR se situe au-dessus du seuil de dégradation de 98% sauf autour de la période du 03/05/2016 où on enregistre une diminution jusqu'à 97.21%. Les causes prépondérantes de l'échec d'établissement RRC sont essentiellement un manque de codes de canalisation, une insuffisance de la puissance en DL ou un manque de contrôleurs de signal au niveau du nodeB.

Nous avons dressé en figure 10, le diagramme du protocole d'acheminement des données dans un système HSDPA dans une autre cellule.

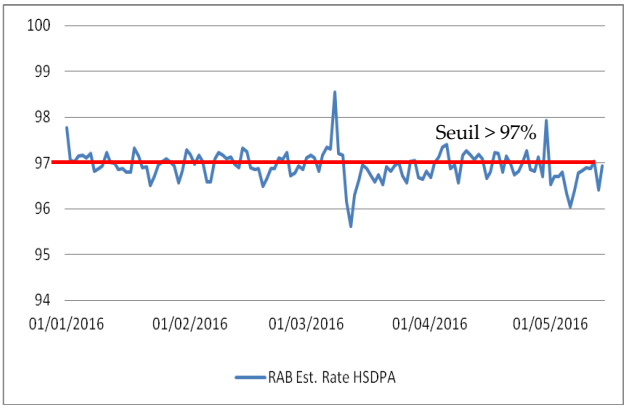


Figure 10. Taux d'acheminement des données dans le domaine Ps [11].

Nous constatons des dégradations permanentes tout au long de la période considérée. La cause essentielle est une congestion causée par un problème de ressources (puissance, codes et transferts cellulaires). Pour y remédier, les ingénieurs radio d'Ooredoo procèdent souvent à un balancement de trafic du réseau 3G vers le réseau GSM ou à un ajustement de consommation des ressources par diminution du débit de façon temporaire.

La figure 11 montre le taux de réussite de soft-handover [7] i.e lors de changements de cellules avec la même fréquence. Les changements de cellule se font avec succès et aucune coupure ne surgit lors de ces changements, donc le soft handover est maitrisé totalement.

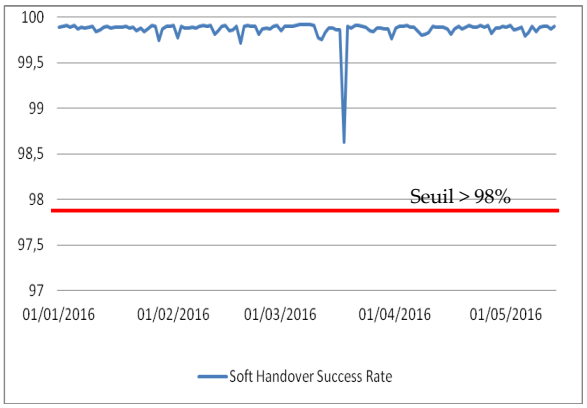


Figure 11. Évolution du taux de réussite de soft handover [11].

3) Réseaux UMTS Mobilis /année 2018

La figure 12 montre le taux du succès du soft handover [7] sur 3 cellules de sites 3G Mobilis à Béjaia en 2018. Cette procédure permet de diminuer le taux d'échec de handover aux bords des cellules. On remarque une bonne maitrise de cette procédure pendant la période d'analyse, même s'il y'avait une légère dégradation pour la cellule (064203Z).

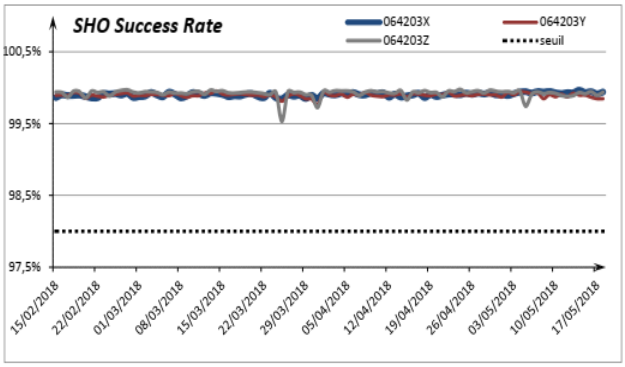


Figure 12. Le taux du succès du Soft handover [12].

La figure 13 montre les résultats du trafic dans le domaine CS [7], pour les trois groupes de cellules (064308U ; 064308V ; 064398W) pendant une durée de 3 mois (de 15/02 à 18/04/18). Pour les mesures de ce KPI, il n'y a pas de seuil fixe, car cela dépend de la taille des cellules ou du RNC. Mais par exemple une valeur très faible révèle un un dysfonctionnement du NodeB.

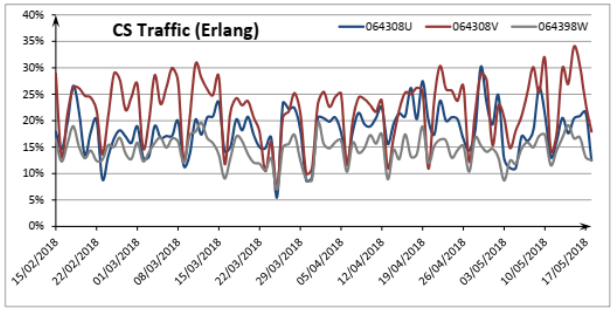


Figure 13. Mesure du trafic CS sur un trimestre 2018 [12].

4) Réseaux 4G Mobilis /année 2018

Une investigation a été menée sur la QoS des sites 4G Mobilis dans quelques à Béjaia en 2018. La Fig.14 illustre l'évolution du taux de coupures de support ERAB [8] sur les sites (06204L, 06664L, 06667L). Pour rappel, un E-RAB est le support de la couche d'accès pour transporter les données de service des utilisateurs.

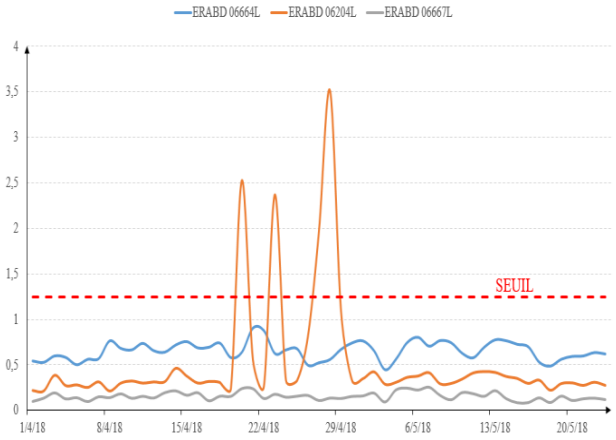


Figure 14. Taux de coupure de support E-RAB sur un mois [13].

On remarque sur les sites (06664L, 06667L) que le taux de coupures est acceptable puisqu'en dessous du seuil (1.25%) cependant il est assez élevé sur le site 06204L du 19 au 29/04/2018. Ces irrégularités peuvent être dues à plusieurs causes notamment l'indisponibilité de débits plus importants demandés par les services, la congestion sur le réseau, la mauvaise couverture, ou la présence d'interférences.

Les figures ci-dessous (Fig.15, Fig.16) illustrent la dégradation des KPIs d'intégrité (débit utilisateur) [8] dans le sens montant et le sens descendant sur les sites (06206L, 06209L, 06211L). On constate de faibles débits (en dessous du seuil 5,4Mps en uplink et 14Mbps en downlink) dans les deux cas de figure. Cette baisse de débit s'explique par des soucis de couverture, des interférences inter sites ou une congestion au niveau la cellule concernée.

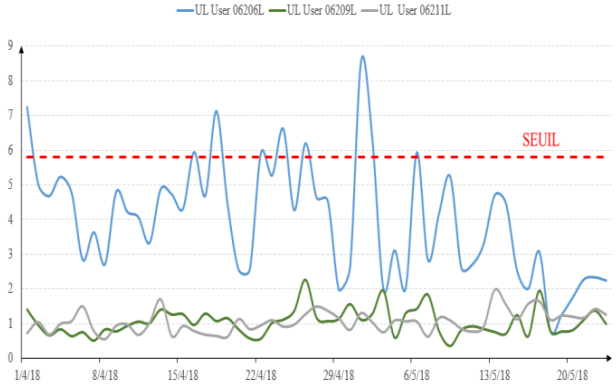


Figure 15. Débit utilisateur en UL sur un mois [13].

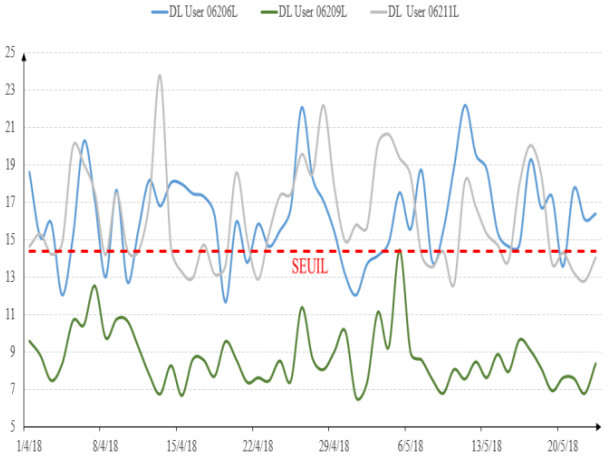


Figure 16. Débit utilisateur en DL sur un mois [13].

5) Réseaux 4G Mobilis /année 2022

Les figures suivantes (Fig.17, Fig.18), montrent le bon fonctionnement des KPIs d'accessibilité (RRC Setup Succ_Rate et S1 Signaling SR) [8] sur 3 cellules (06686 N, 06686M, 06686O) durant un semestre 2022.

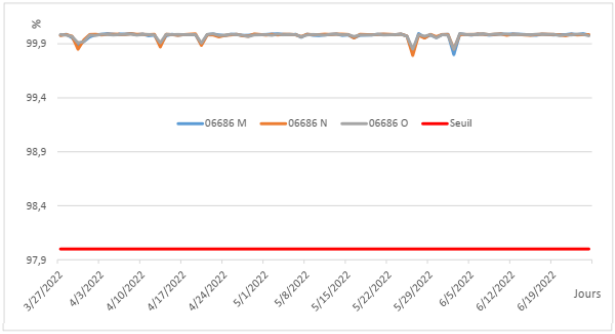


Figure 17. Taux de succès d'établissement RRC [14].

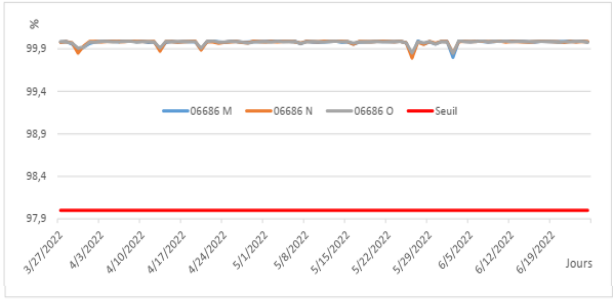


Figure 18. Taux de succès de signalisation S1[14].

6) Réseaux 4G Mobilis /année 2023

Les figures suivantes, (Fig.19, Fig.20), montrent le taux d'échec d'établissement de la connexion RRC CS pour les connexions vocales et le taux de réussite IRAT CS (Inter-Radio Access Technology Circuit-Switched) [8] mesurant la réussite des transferts de connexion entre différentes technologies d'accès radio, telles que la transition entre les réseaux circuit-switched 2G (GSM) et 3G (UMTS) sur des cellules à Béjaia durant un semestre 2023.

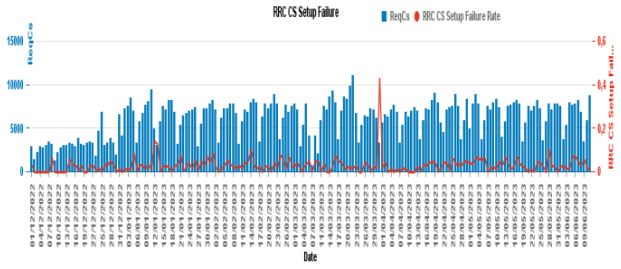


Figure 19. Taux d'échec RRC CS Setup [15].

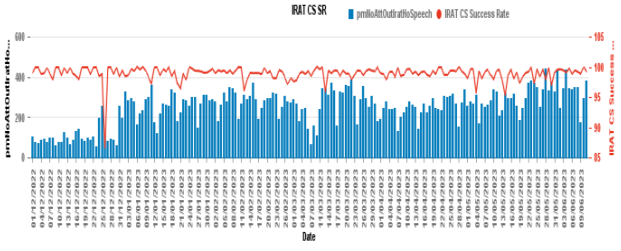


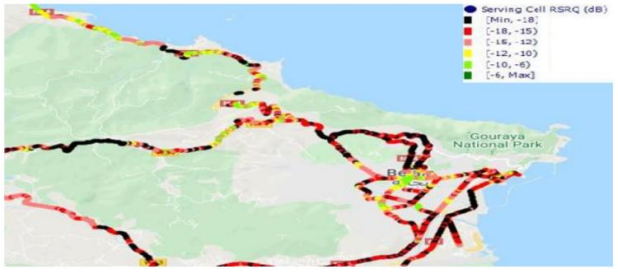
Figure 20. IRAT CS Success Rate [15].

B. Optimisation du réseau radio Mobilis après drive test

Les figures 21 et 22, donnent les résultats d'analyse selon le niveau de puissance de réception des signaux de références RSRP, et selon la la qualité des signaux de référence à la réception RSRQ.

Couleur							
RSRP (dBm)	≥-75	-75 to -85	-85 to -95	-95to-110	-110 to115	≤-115	Good RSRP ≥-95
Proportion	4,42 %	12,83 %	23,82 %	28,35 %	17,50 %	13,09 %	41,06 %
Moyenne	-98,52 dBm						

Figure 21. Parcours et distribution RSRP lors du DT [16].



RSRQ (dB)	≥-6	-6 to-10	-10to-12	-12to-15	-15to-18	<-18	Good RSRQ ≥-12 (dB)
Proportion	0,17 %	7,99 %	13,29 %	24,77 %	25,36 %	28,42 %	21,45 %
Moyenne	-15,72 dB						

Figure 22. Parcours et distribution RSRQ lors du DT [16].

Les valeurs disparates du RSRP, révèlent une distribution inégale de la couverture sur la zone objet du drive-test. Bien que 41% de la zone bénéficie d'une bonne couverture, une proportion significative de 31% a une mauvaise couverture et une autre de 28% une couverture moyenne.

De plus, les valeurs RSRQ mesurées montrent qu'une proportion de 54% de la zone souffre d'une mauvaise qualité, synonyme ici de présence d'interférences et d'un bruit élevé pouvant entraîner une connectivité instable et une détérioration de la qualité des communications.

Ces résultats ont conduit à une densification du réseau sur le plan de couverture par un ajout de nouveaux sites dans la zone étudiée.

Un autre cas d'optimisation a été traité suite à la constatation de baisse de performances du site 23 de Mobilis desservant le campus targua ouzemmour. En effet les figures suivantes (Fig.23, Fig.24) montrent que malgré des proportions élevées de PRB utilisés, les débits en DL sont relativement faibles dans les 3 cellules du site, ce qui se traduit par une efficacité limitée dans la transmission des données des usagers. La solution adoptée a été d'implémenter une technologie massive MIMO.

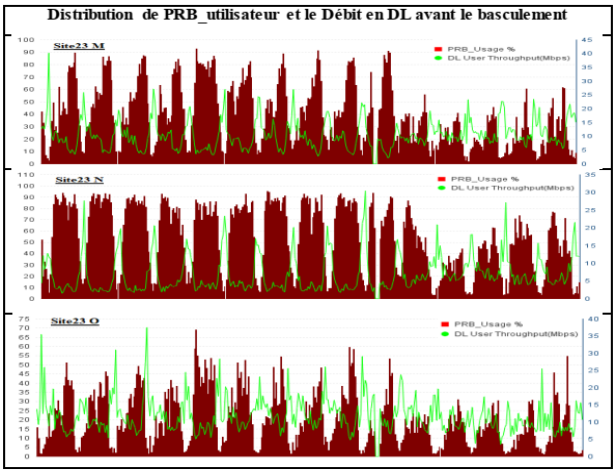


Fig. 23. Valeurs antérieures des PRB_users et du Débit DL (site23) [16].

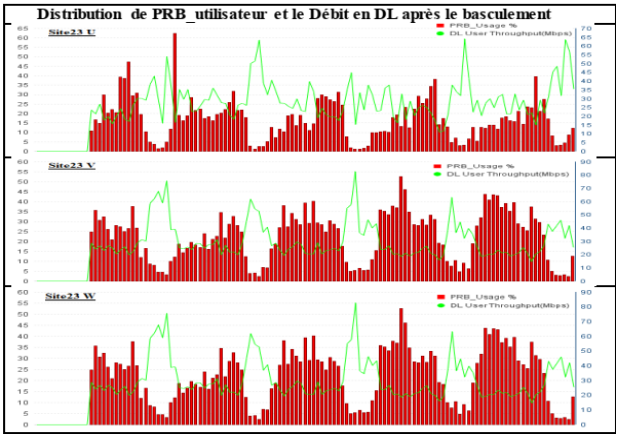


Figure 24. Valeurs postérieures au basculement en massive MIMO [16].

C. *Analyse statistique d'alarmes des sites Mobilis à Béjaia*

Cette section reprend le traitement de fichiers d'alarmes 4T sur une durée de 5 mois (janvier à Mai) de l'année en cours sur tous les sites de la région de Béjaia (753). Nous donnons ci-après une analyse statistique des données recueillies dans le but d'améliorer la disponibilité du réseau ATM et d'améliorer la politique de maintenance préventive de ses sites.

Ce tableau présente la variation mensuelle des pannes internes sur la période de suivi.

Tableau 1. Caractérisation mensuelle des pannes des sites ATM [15].

mois / éléments	Janvier	Février	Mars	Avril	Mai
Nombre de panne	221	367	282	619	347
Temps d'arrêt total (h:min:s)	672:51:00	1405:29:00	778:55:00	311:29:00	2194:37:00
Temps d'arrêt (h:min:s)	394:23:00	681:21:00	424:02:00	712:28:00	1613:49:00
MTTR (h:min:s)	1:39:41	3:08:39	2:10:33	1:54:03	4:13:43
Fréquence de panne(%)	31.26%	51,91%	39,89%	87,55%	49,08%

Les figures 24 et 25 donnent une représentation de l'occurrence des pannes et leurs types sur la période de mesure ainsi que le temps d'arrêt (down time) et le temps moyen de réparation (MTTR).

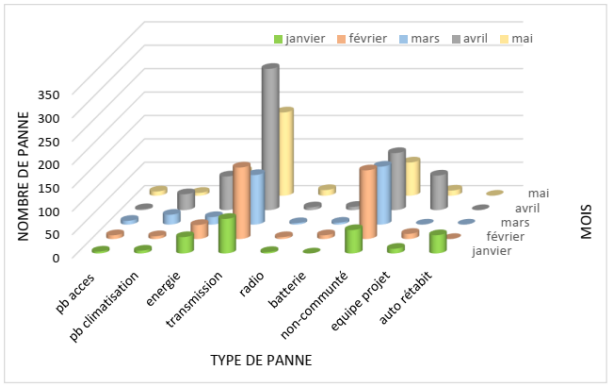


Figure 25. Nombres et types de pannes des sites ATM (S1 2023) [15].

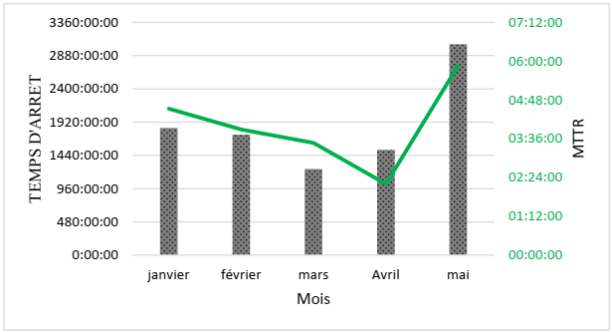


Figure 26. Temps d'arrêt de temps de réparation MTTR des sites [15].

Les différentes mesures nous ont permis de déterminer la disponibilité des 753 sites ATMobilis à Béjaia sur les 5 mois de notre investigation, présentée en Fig.27.

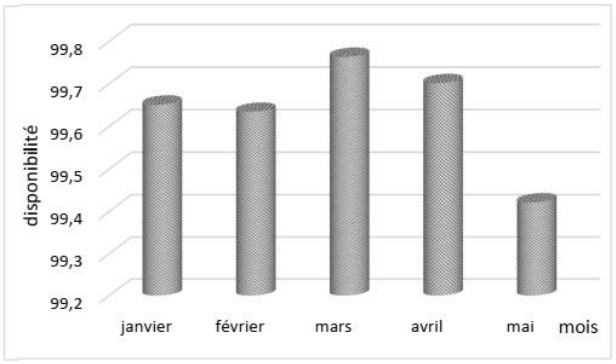


Figure 27. Détermination de la disponibilité des sites ATM [15].

L'idée derrière cette approche est d'identifier les signes précurseurs d'une panne imminente ou les éléments susceptibles de causer des problèmes à l'avenir. En intervenant en amont, avant que la panne ne survienne effectivement, on évite les interruptions de service ou les dysfonctionnements qui pourraient avoir un impact négatif sur les utilisateurs ou les opérations. La maintenance préventive est donc une pratique proactive visant à anticiper et à prévenir les pannes avant qu'elles ne se produisent. Cela peut inclure des inspections régulières, des vérifications des performances, des remplacements préventifs de composants usés ou défectueux, des mises à jour logicielles, etc. En effet, les entreprises cherchent à minimiser les temps d'arrêt, à améliorer la disponibilité du système et à optimiser la performance de leurs équipements. Cela contribue également à réduire les coûts liés aux réparations d'urgence et aux perturbations opérationnelles.

V. CONCLUSION

Cette contribution est une revue de travaux réalisés sur des éléments de supervision, de fiabilité et d'optimisation des performances des réseaux radio-mobiles déployés dans la région de Béjaia. Les résultats et les constats présentés témoignent d'une maitrise de ces différents aspects par les équipes d'ingénieurs des centres OMC des opérateurs présents dans notre localité.

REFERENCES

[1] Saad Z. Asif, "5G Mobile Communications: Concepts and Technologies", CRC Press, Taylor & Francis Group, 2019.

[2] Martin Sauter, "From GSM to LTE-Advanced Pro and 5G : An Introduction to Mobile Networks and Mobile Broadband", 3rd edition, John Wiley & Sons Ltd, 2017.

[3] F. Launav. "NG-RAN et 5G-NR: L'accès radio 5G et l'interface radioélectrique", ISTE Group, 2021

[4] X. Lagrange, P. Godlewski, S. Tabbane, "Réseaux GSM : des principes à la norme ", Hermès Science Publications, 2000.

[5] Y. Ouazziz,, H. Messaoudi, "Eléments de supervision & d'optimisation de réseaux radio mobiles", mémoire Master 2 en réseaux et télécommunications, UAMBéjaia, 2019.

[6] Ericsson Confidential ; GSM Network KPI Root Cause analysis.

[7] Ericsson WCDMA W14 Radio Network functionality.

[8] Ericsson LTE KPIs Description.

[9] A. Mishra, "Fundamentals of Network Planning and Optimisation 2G/3G/4G", 2nd Edition, Wiley, 2018.

[10] T. Kaci, K. Ferhane, « Étude de la qualité de service du réseau GSM/Mobilis à Bejaia », mémoire Master télécommunications, UAMBéjaia, 2015.

[11] K.Bouhaddi, D. Benhamla, "Evaluation de la qualité de service voix et données du réseau 3G Ooredoo à Bejaia", mémoire Master télécommunications, UAMBéjaia, 2016.

[12] S. Issaad, H. Kahouadji, "Analyse et Optimisation d'Indicateurs de QoS du réseau 3G/WCDMA d'AT Mobilis de Bejaia », mémoire Master 2 en Systèmes de télécommunications, UAMBéjaia, 2018.

[13] I.O. Touré, W. Diarra, "Analyse et optimisation des indicateurs QoS de réseaux 4G/LTE". mémoire Master 2 en réseaux et télécommunications, UAMBéjaia, 2018.

[14] K. Boumeridja, Z. Ouari, "Évaluation de la fiabilité et de la QoS de réseaux radio mobiles". mémoire Master 2 en réseaux et télécommunications, UAMBéjaia, 2022.

[15] L. Amia, F. Barache, "Gestion des alarmes pour l'optimisation de la disponibilité des sites radio mobiles à Bejaia". mémoire Master 2 en réseaux et télécommunications, UAMBéjaia, 2023.

[16] C. Megrous, K. Hamma, "Densification et paramétrage des sites Radio mobiles 4G à Bejaia ", mémoire Master 2 en réseaux et télécommunications, UAMBéjaia, 2023.

IV

MODÈLES STOCHASTIQUES

Sommaire

Reliability analysis of a repairable redundant system with unreliable repairer

Boudehane Kheireddine
University of Science and Technology Houari Boumediene, RIIMA laboratory
k_boudehane@yahoo.com; kboudehane@usthb.dz

Taleb Samira
University of Science and Technology Houari Boumediene, RIIMA laboratory
talebsamira04@yahoo.fr

***Abstract*—**
This paper studies the reliability and availability of a $K-out-of-M+W+C:G$ retrial system with a single unreliable repairer. The system is modeled using a Generalized Stochastic Petri Net ($GSPN$). The repairer can fail in two ways: while it is repairing a component (active breakdown) or while it is idle (passive breakdown). The failure times of the repairer and the components are all assumed to be exponentially distributed. When a component fails, it is either repaired immediately if the repairer is available and free, or it enters a retrial orbit if the repairer is busy or under repair. We use the $GSPN$ model to derive a Continuous Time Markov Chain ($CTMC$), which allows us to calculate the system's main stationary probabilities and some performance reliability measures.

***Keywords*—**
Generalized stochastic Petri net, $K-out-of-n$ system, reliability, availability.

I. Introduction

A $K-out-of-n$ system is a system that consists of n components, and the system is considered to be successful if at least K of the n components are functioning properly. The value of K must be less than or equal to n.

$K-out-of-n$ systems are often used in fault-tolerant systems. In a fault-tolerant system, the goal is to ensure that the system continues to function even if some of the components fail. $K-out-of-n$ systems can be used to achieve this goal by providing redundancy. $K-out-of-n$ systems can be analyzed using a variety of methods, including:

- Reliability analysis: This method uses the mathematical theory of reliability to calculate the probability of the system being successful.
- Fault tree analysis: This method uses a graphical representation of the system to identify and analyze potential failure modes.
- Simulation: This method uses a computer program to execute the system model and generate a sample path of the system behavior.

Standby redundancy is a fault-tolerant system design that ensures high availability and reliability by having backup components or systems on standby, ready to take over in case the primary component or system fails. There are three main types of standbys: hot, cold, and warm.

- Hot standby: a component is in hot standby if its failure rate in standby mode is the same as its failure rate in active mode.
- Cold standby: A component is in cold standby if its failure rate in standby mode is zero, meaning that it cannot fail while in standby.
- Warm standby: A component is in warm standby if its failure rate in standby mode falls somewhere in between hot and cold standby.

Generalized Stochastic Petri Nets ($GSPNs$) [1], are Petri nets (PNs) that have a probabilistic timing mechanism, used to model the behavior of complex systems. A $GSPN$ comprises:

- Places: represent conditions or objects in the system. They are represented by circles in a $GSPN$ diagram.
- Tokens: represent the presence of an object or condition in a place. They are represented by black dots or numbers in a $GSPN$ diagram.
- Transitions: represent events in the system. They are represented by rectangles in a $GSPN$ diagram, they are partitioned into :
  - Immediate transitions: fire instantaneously, without any delay. They are represented by thin bars in a $GSPN$ diagram.
  - Timed transitions: have a delay before they can fire. The delay is exponentially distributed. They are represented by rectangles in a $GSPN$ diagram.
- Arcs: connect places and transitions. They indicate the flow of tokens between places and transitions.
- Inhibitor arcs: that prevent a transition from firing.

Enabled transition: An enabled transition in a $GSPN$ refers to a transition whose input places have enough tokens to allow the transition to be fired. If the firing transition is timed, the marking is referred to as tangible marking and if the firing transition is immediate, the marking is referred to as vanishing markings.

Reachability graph: A reachability graph of a $GSPN$ is a directed graph that represents all of the possible markings (states) of the net. The nodes of the graph

represent the markings of the net, and the edges of the graph represent the firing of transitions. The tangible reachability graph in a $GSPN$ has a isomorphic with a Continuous Time Markov Chain ($CTMC$).

Live $GSPN$: A live $GSPN$ is a $GSPN$ in which every transition is enabled from some reachable marking. This means that it is possible to fire any transition in the net, regardless of the current state of the system. Liveness is an important property of Petri nets, as it ensures that the net cannot become deadlocked. A deadlock is a state in which no further transitions can be fired, and the system is effectively stuck. There are a number of ways to check if a $GSPN$ is live. One common approach is to use a reachability graph. By traversing the reachability graph, it is possible to identify any states from which no further transitions can be fired. If the reachability graph contains no deadlocks, then the $GSPN$ is live.

Home State: A home state of a $GSPN$ is a marking that is reachable from every reachable marking. In other words, it is a marking to which the system may always return.

Bounded $GSPN$: A bounded Generalized Stochastic Petri Net (GSPN) is a GSPN in which the number of tokens in each place is always bounded by a finite number.

Steady-state probability distribution of $GSPN$: If a Generalized Stochastic Petri Net (GSPN) is bounded and its initial marking is a home state, then its steady-state probability distribution exists.

$GSPNs$ can be used to model a wide variety of systems, including:

- Computer systems [2]: $GSPNs$ can be used to model the behavior of computer systems, such as the flow of control in a program.
- Manufacturing systems [3]: $GSPNs$ can be used to model the behavior of manufacturing systems, such as the flow of materials through a production line or the behavior of a robotic assembly system.
- Business processes [4]: $GSPNs$ can be used to model the behavior of business processes, such as the flow of work through an organization or the behavior of a customer service system.

$GSPNs$ can be analyzed using a variety of methods, including:

- Simulation: This is the most common method for analyzing $GSPNs$. A simulation is a computer program that executes the $GSPN$ model and generates a sample path of the system behavior.
- Markov analysis: This method uses the mathematical theory of Markov chains to analyze the $GSPN$ model. Markov analysis can be used to calculate the probability of different events occurring in the system, such as the probability of a system failure or the average time it takes for a task to be completed In this paper, we study a $K-out-of-M+W+C:G$ retrial system with unreliable repairer, through a $GSPN$.

II. Model description

The system consists of M primary operational components, W warm standby components, and C cold standby components. If a primary operating component fails, an available warm standby component immediately takes its place. If a warm standby component fails during the standby period, an available cold standby component immediately takes its place or replaces a primary component. The repairer may experience a failure during the repair period (active breakdowns) or during a period of inactivity (passive breakdowns). The time needed for a repairer to recover and resume its service is characterized by an exponential distribution. If a new failed component finds the repairer busy or under repair, it enters an orbit for a random period and waits for repair.

The lifetimes of primary and warm standby components follow exponential distributions with parameters λ and α ($\lambda > \alpha > 0$) respectively. When a component fails, it will be repaired immediately if the repairer is available. Repair time for any component follows an exponential distribution with parameter μ.

The lifetime of the repairer is assumed to be exponentially distributed with rate δ when idle and θ when busy. The time needed for the repairer to recover and resume its service is characterized by an exponential distribution with rate σ.

When a new component fails and finds the repairer is either busy or out of order, it enters an $FCFS$ orbit for a random amount of time until the repairer becomes available. Once the repairer becomes available again, the selection time for a failed component from the orbit, if any, follows an exponential distribution with a mean of $\frac{1}{\gamma}$.

Lifetimes, repair times, and retrial times are assumed to be independent.

III. $GSPN$ model for $K-OUT-OF-M+W+C:G$ retrial system with unreliable repairer

A. GSPN description

The $GSPN$ model for a $K-out-of-M+W+C:G$ retrial system with a single unreliable repairer is shown in Figure 1. The model has 9 places, 7 timed transitions, and 4 immediate transitions. The places in the model are as follows:

- P_M: primary operating components, with initial marking M,
- P_W: warm standby components, with initial marking W,
- P_C: cold standby components, with initial marking C,

- $Choice$ condition that a new failed component or a repeated component is ready for repair,
- $Orbit$:represents Orbit,
- P_{Rep}: failed components in repair,
- $P_{R.fail}$: failed repairer,
- $P - R$: the free repairer,
- P_L: number of failed components.

The initial marking of the $GSPN$ is given by $M_0 = (M, W, C, 0, 0, 1, 0, 0, 0)$
The timed transitions in the model are as follows:

- t_λ: fail of a primary operating component with rate λ,
- t: fail of a warm component with rate μ,
- t_γ: arrival of a repeated component from orbit with rate γ,
- t_μ: end of repair period with rate μ,
- t_δ: failure of a repairer during a repair period with rate δ,
- t_θ: failure of a repairer during an idle period with rate θ,
- t_σ: end of repair time for a failed repairer with rate σ. The immediate transitions in the model are as follows:
- t_{Rep}: fires if there is a failed of primary or repeated call in place $Choice$, and place P_R contains one token (the repairer is available),
- t_O: fires if the repairer is busy or under repair, and the failed component is immediately placed in Orbit, and waits for repair,
- t_W: fires if place P_M contains tokens less than M, that consists of destroying one token from place P_W and constructing a token in the place P_M which means that available warm component replaces failed primary component,
- T_C: fires if place P_W contains tokens less than W, that consists of destroying one token from place P_C and constructing a token in the place P_W which means that available cold component replaces failed warm standby components.

The system state at time t can be described by (i, j, k) where :

- i: The number of failed componentsin repair,
- j: The number of failed components in orbit,
- $k = 1$ if the repairer is failed, and $k = 0$ if the repairer is free and availaible

IV. Example

We fixed the values of M, W, C, and K to 2, 2, 1, and 2, respectively. We then obtained the tangible reachability tree, which describes all possible states of our $GSPN$ starting from the initial marking $M_0 = (0, 0, 0)$. From the tangible reachability tree, we obtained the state-transition-diagram of the $GSPN$, which is shown in Figure 2. This diagram shows all possible transitions between states of the $GSPN$.

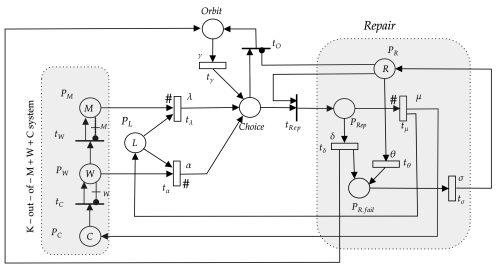


Figure 1. $GSPN$ model for $K - out - of - M + W + C : G$ retrial system with single unreliable repairer

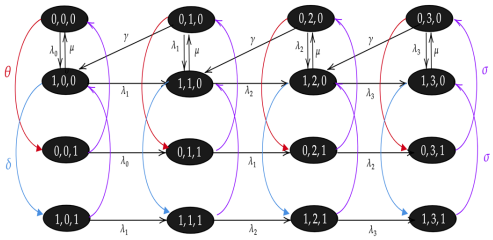


Figure 2. Diagram-state of $CTMC$ derived from the $GSPN$ model

The obtained $GSPN$ model of Figure1 is bounded and the initial marking is a home states then their steady-state probability distributions exist. After obtaining the row vector of the stationary probabilities $\pi_{i,j,k}, 0 \leq i \leq 1, 0 \leq j \leq L$ and $0 \leq k \leq 1$, we can derive some performance and reliability measures of the $K - out - of - M + W + C : G$ retrial system.

- The steady-state availability:

$$A(\infty) = 1 - (\pi_{1,3,0} + \pi_{1,3,1}) \quad (1)$$

- Mean number of failed components in repair:

$$N_r = \sum_{j=0}^{4} (\pi_{1,j,0} + \pi_{1,j,1}) \quad (2)$$

- Mean number of failed components in orbit:

$$N_0 = \sum_{j=0}^{4} (\pi_{0,j,0} + \pi_{0,j,1} + \pi_{1,j,0} + \pi_{1,j,1}) \quad (3)$$

- Reliability function:

$$R(t) = 1 - (\pi_{1,3,0}(t) + \pi_{1,3,1}(t)) \quad (4)$$

V. NUMERICAL RESULTS

In this section we present some numerical results. We take $\lambda = 0.6$, $\alpha = 0.05$, $\mu = 2$, $\gamma = 3$, $\sigma = 1$, $\theta = 0.5$ and $\delta = 0.8$ as the base case. To explain the effect of each system parameter on $R(t)$ and $A\infty$, we change the values of parameters (λ, μ and γ) in turn while other parameters take the given values in the base case. The numerical results of $A\infty$ and $R(t)$ are shown in Figures 3-4 and Figures 5-7, respectively. Next, we present the effect of λ and μ on Mean number of failed components in repair N_r and Mean number of failed components in orbit N_O (see Tables I-II)

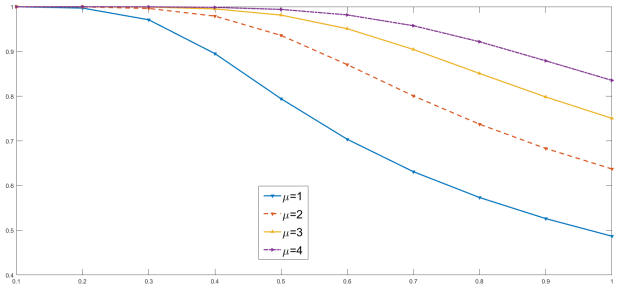


Figure 3. A(∞) versus λ for different values of μ

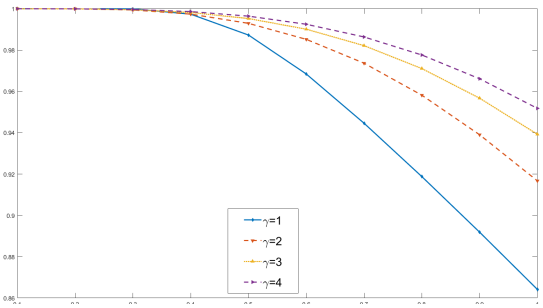


Figure 4. A(∞) versus λ for different values of γ

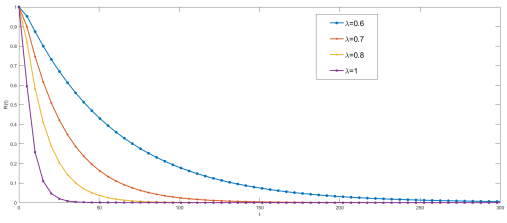


Figure 5. $R(t)$ versus t for different values of λ

VI. CONCLUSION

This paper proposes a method for evaluating the performance and reliability of a complex system called a $K - out - of - M + W + C : G$ retrial system with warm and cold standbys and unreliable repairer. The method is based on Generalized Stochastic Petri Nets ($GSPNs$). We first derive a continuous-time Markov chain ($CTMC$) with finite state space from the $GSPN$ model. The $CTMC$ is then solved to obtain the system's steady-state probabilities, the steady-state availability, and some performance measures. Finally, it is important to note that our approach can be extended to even more complex systems.

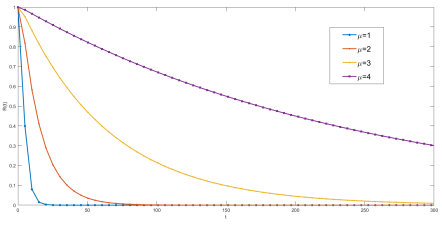


Figure 6. $R(t)$ versus t for different values of μ

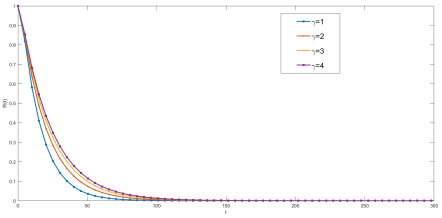


Figure 7. $R(t)$ versus t for different values of γ

REFERENCES

[1] M. Ajmone Marsan, G. Balbo, G. Conte, S. Donatelli, G. *"Franceschinis, Modelling with Generalized Stochastic Petri Nets"*, New York: John Wiley Sons, 1995

Table I. Performance measures N_r and N_o for different values of λ

Measures	N_r	N_O
$\lambda = 0.2$	0.015	2.988
$\lambda = 0.3$	0.019	3.249
$\lambda = 0.4$	0.021	3.432
$\lambda = 0.5$	0.080	3.566
$\lambda = 0.6$	0.127	3.668
$\lambda = 0.7$	0.165	3.566
$\lambda = 0.8$	0.197	3.811
$\lambda = 0.9$	0.224	3.862
$\lambda = 1$	0.249	3.903

Table II. Performance measures N_r and N_o for different values of μ

Measures	N_r	N_O
$\mu = 1$	0.88	3.857
$\mu = 2$	0.127	3.668
$\mu = 3$	0.098	3.620
$\mu = 4$	0.078	3.597
$\mu = 5$	0.064	3.583

[2] N.Gharbi, and L.Charabi, "Wireless Networks with Retrials and Heterogeneous Servers:Comparing Random Server and Fastest Free Server Disciplines", "*International Journal on Advances in Networks and Services*", 2012, 5, pp 102-115.
[3] S. Viswanadham, Y. narahari and L. Jhonson, "Deadlock Prevention and Deadlock Avoidance in Flexible Manufacturing Systems Using Petri Net Models", "*IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION*", 1990, 6(6).
[4] A.L. Cesar, M.F.Lima : "Performance Analysis of Resource-Constrained Business Processes: a Formal Approach Based on Stochastic Petri Nets", "*www.cin.ufpe.br*", Brazil, 2009.

Fuzzy retrial queue with breakdowns and repairs.

Boussaha Zina
Military Academy of Cherchell DPHB, ALGERIA.
boussaha_z@yahoo.fr

Oukid Nadia
LAMDA-RO Saad Dahlab University, Blida 1, ALGERIA.
oukidnad@yahoo.fr

***Abstract*—**
The present work uses the flexible α-cuts method to analyse a retrial $M/M/1$ queue with breakdowns and repairs in fuzzy environment. Through the illustrative example proposed below, we show that the flexible alpha-cuts method is suitable not only for fuzzy queues having at most three parameters, but also for those contain at most five parameters.

***Keywords*—**
fuzzy queue, the flexible α-cuts method, Zadeh's extension principle, breakdowns, repairs.

I. Introduction

Queueing system with repeated attempts have wide practical use in designing telephone switching systems, telecommunication networks, computer networks and computer systems, etc. Since in practice some components of the systems are subject to random breakdowns it is of basic importance to study reliability of retrial queues with server breakdowns and repairs because of limited ability of repairs and heavy influence of the breakdowns on the performance measures of the system. However, so far the repairable retrial queues are analyzed only by queueing theory. A review of main results of the crisp model can be found in [1], [4], [10].

Regarding the fuzzy model, only a few studies investigated this topic. Amoung them, [2] and [9] combined Zadeh's extension principle and mathematical nonlinear programming method to compute performance measures. Using L-R method (Left-Right method) introduced by [7], [5] computed performance measures of the $M/M/1$ fuzzy model. In [6] the author used a new technique introduced recently in [8], called *"flexible α-cuts method"* to analyse the $M/M/1$ model in fuzzy environment. Let us recall that this technique consists in applying the alpha-cut and interval arithmetic to fuzzy queueing formulae without resorting to other forms of computation. Its flexibilty and advantage compared to other methods are in the fact that nearly all calculations are nonfuzzy. To show its powerful, we apply it here to queueing model with multimodal fuzzy parameters instead of unimodal fuzzy parameters applied in [8]. Through the illustrative example proposed below, we show that the flexible alpha-cuts method is suitable not only for fuzzy queues having at most three parameters, but also for those contain at most five parameters.

II. Preliminaries

A. Fuzzy set

- Let E be a classical set or a universe. A fuzzy subset $\tilde{A}$ in E is defined by the function $\eta_{\tilde{A}}$, called membership function of $\tilde{A}$, from E to the real unit interval [0, 1].
- The alpha-cut $\widetilde{A}_\alpha$, the support $supp(\tilde{A})$ the height $hgt(\tilde{A})$, and the core $core(\tilde{A})$ of $\tilde{A}$, are crisp sets defined respectively as follows $\forall\alpha \in$ [0, 1].

$$\widetilde{A}_\alpha = \left\{x \in E \mid \eta_{\widetilde{A}}(x) \geq \alpha\right\}, \tag{1}$$

$$\mathrm{supp}(\widetilde{A}) = \left\{x \in E \mid \eta_{\widetilde{A}}(x) > 0\right\}, \tag{2}$$

$$\mathrm{hgt}(\widetilde{A}) = \sup\left\{\eta_{\widetilde{A}}(x) \mid x \in E\right\}, \tag{3}$$

$$\mathrm{core}(\widetilde{A}) = \left\{x \in E \mid \eta_{\widetilde{A}}(x) = 1\right\}. \tag{4}$$

The membership function of a fuzzy set $\tilde{A}$ can be expressed in terms of characteristic functions of its α-cuts according to the formula

$$\eta_{\widetilde{A}}(x) = \sup_{\alpha\in[0,1]} \min\left\{\alpha, \eta_{\widetilde{A_\alpha}}(x)\right\} \tag{5}$$

where

$$\eta_{\widetilde{A}_\alpha}(x) = \begin{cases} 1 & \text{if } x \in \widetilde{A}_\alpha, \\ 0 & \text{otherwise.} \end{cases}$$

- A fuzzy set $\tilde{A}$ is said normal if and only if $\mathrm{hgt}(\widetilde{A}) = 1$, and $convex$ if and only if $\eta_{\widetilde{A}}(\lambda x + (1-\lambda y)) \geq \min\{\eta_{\widetilde{A}}(x), \eta_{\widetilde{A}}(y)\}, \forall x, y \in E, \forall\alpha \in [0,1]$.

B. Fuzzy number

- A fuzzy set $\tilde{A}$ is called a *fuzzy number* if $\tilde{A}$ is a fuzzy subset of $\mathbb{R}$ such as:
  1) $\mathrm{core}(\widetilde{A}) = \emptyset$;
  2) $\widetilde{A}_\alpha$ are all closed and bounded subintervals of $\mathbb{R}$;
  3) $\mathrm{supp}(\widetilde{A})$ is bounded.
- A fuzzy set $\tilde{A}$ is called a *strictly positive* if $\eta_{\widetilde{A}_\alpha}(x) = 0, \forall x < 0$ and *strictly negative* if $\eta_{\widetilde{A}_\alpha}(x) = 0, \forall x > 0$.
- Let $\tilde{A}$ and $\tilde{B}$ be two fuzzy numbers:

$$\widetilde{A} < \widetilde{B} \Leftrightarrow \sup\{\mathrm{supp}(\widetilde{A})\} < \inf\{\mathrm{supp}(\widetilde{B})\}. \tag{6}$$

- A fuzzy set $\tilde{A}$ is called a *trapezoidal fuzzy number*, noted $\tilde{A} = (a/b/c/d)$, if and only if there is four real numbers $a < b < c < d$ such that:

$$\eta_{\widetilde{A}}(x) = \begin{cases} \frac{x-a}{b-a} & \text{if} \quad a \leq x \leq b, \\ 1 & \text{if} \quad b < x \leq c, \\ \frac{d-x}{d-c} & \text{if} \quad c < x \leq d, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Remark: Trapezoidal fuzzy numbers belong to the set of fuzzy numbers whose the core contains at least two elements, called multimodal fuzzy numbers or flat fuzzy numbers in [3].

III. Arithmetic of intervals and α-cuts

- Let $[a,b],[c,d]$ two bounded real intervals and $*$ the classical operation of addition, subtraction, multiplication or division. We have :

$$[a,b] * [c,d] = [\alpha,\beta], \tag{8}$$

where $[\alpha,\beta] = \{x * \frac{y}{a} \leq x \leq b, c \leq y \leq d\}$ assuming that $0 \notin [c,d]$ for the division.
In concrete terms, we have :

$$[a,b] + [c,d] = [a+c, b+d] \tag{9}$$

$$[a,b] - [c,d] = [a-d, b-c] \tag{10}$$

$$[a,b] \times [c,d] = [\min\{ac,ad,bc,bd\}, \max\{ac,ad,bc,bd\}] \tag{11}$$

$$[a,b] \div [c,d] = \left[\min\left\{\frac{a}{c},\frac{a}{d},\frac{b}{c},\frac{b}{d}\right\}, \max\left\{\frac{a}{c},\frac{a}{d},\frac{b}{c},\frac{b}{d}\right\}\right] \tag{12}$$

- Let $\tilde{A}$, $\tilde{B}$ be two fuzzy numbers of respective alpha-cuts $\tilde{A}_\alpha = \left[A^L(\alpha), A^U(\alpha)\right]$ and $\tilde{B}_\alpha = \left[B^L(\alpha), B^U(\alpha)\right]$. The fuzzy arithmetic operations on $\tilde{A}$ and $\tilde{B}$ are defined via their α-cuts in the following way:

$$[\tilde{A} \oplus \tilde{B}]_\alpha = \tilde{A}_\alpha + \tilde{B}_\alpha = \left[A^L(\alpha), A^U(\alpha)\right] + \left[B^L(\alpha), B^U(\alpha)\right] \tag{13}$$

$$[\tilde{A} \Theta \tilde{B}]_\alpha = \tilde{A}_\alpha - \tilde{B}_\alpha = \left[A^L(\alpha), A^U(\alpha)\right] - \left[B^L(\alpha), B^U(\alpha)\right] \tag{14}$$

$$[\tilde{A} \otimes \tilde{B}]_\alpha = \tilde{A}_\alpha \times \tilde{B}_\alpha = \left[A^L(\alpha), A^U(\alpha)\right] \times \left[B^L(\alpha), B^U(\alpha)\right] \tag{15}$$

$$[\tilde{A} \oslash \tilde{B}]_\alpha = \tilde{A}_\alpha \div \tilde{B}_\alpha = \left[A^L(\alpha), A^U(\alpha)\right] \div \left[B^L(\alpha), B^U(\alpha)\right] \tag{16}$$

The α-cuts in Equations (13), (14), (15), and (16) are often computed by using intervals arithmetic formulas in Equations (9), (10), (11), and (12).
Remark: To effect fuzzy arithmetic by " α-cuts and intervals arithmetic" consists in using successively:
  - Equations (13), (14), (15), and (16) for defuzzification.
  - Equations (9), (10), (11), and (12) for classical arithmetic on real closed intervals.
  - Equation (5) for fuzzification.

IV. Description of the model

We consider a single server retrial queue with orbital search of customers. The detailed description of the model is given as follows:

The arrival process: Customers arrive at the system according to a Poisson process with fuzzy parameter $\tilde{\lambda}$.

The service process: Service time of incoming customers follows the exponential distribution with fuzzy rate $\tilde{\gamma}$.

The failure process: The server is subject to breakdowns whose failure times are independent and exponentially distributed with fuzzy rate $\tilde{\rho}$.

The retrial process: Once a customer is interrupted by server breakdown, he leaves the service area and enters into the retrial orbit whose rate is a fuzzy number $\tilde{\theta}$. Orbit customers do not rejoin the normal queue but rather attempt to access the server directly after an uncertain amount of time when the server is again operational and idle.

The repair process: The server is assigned in repair whose rate is a fuzzy number $\tilde{\beta}$, when the customer rejoined the orbit.

All processes in the system are assumed to be independent and identically distributed. The queue size and the orbit size are assumed to be infinite and the service discipline is FIFO (first in first out).

A. Stability condition in fuzzy model

For the crisp model, [9] gives the following system stability condition:

$$\frac{\lambda(\beta+\rho)}{\beta\gamma} < 1, \tag{17}$$

and for the analytical crisp formula for the customers number in the queue noted N_q, we have:

$$N_q = \frac{\lambda\left[\gamma\sigma(\gamma+\sigma) + \lambda(\beta+\sigma)^2\right]}{\gamma(\beta+\sigma)[\beta(\gamma+\sigma) - \lambda(\beta+\sigma)]}. \tag{18}$$

for computations in fuzzy model, it is important to verify the queue stability condition. For (6) and (17) we have:

$$\sup\{\text{supp}[\widetilde{\lambda}(\widetilde{\beta}+\widetilde{\rho})]\} < \inf\{\text{supp}(\tilde{\gamma}\tilde{\beta})\}, \tag{19}$$

and we get:

$$\tilde{N}_q = \frac{\tilde{\lambda}\left[\tilde{\gamma}\tilde{\sigma}(\tilde{\gamma}+\tilde{\sigma}) + \tilde{\lambda}(\widetilde{\beta}+\tilde{\sigma})^2\right]}{\tilde{\gamma}(\widetilde{\beta}+\tilde{\sigma})[\widetilde{\beta}(\tilde{\gamma}+\tilde{\sigma}) - \tilde{\lambda}(\widetilde{\beta}+\tilde{\sigma})]}, \tag{20}$$

V. Computation process of flexible α-cuts method

Let us precise as in the introduction that the present process can be applied in fuzzy model only when the analytical crisp queue formula and the fuzzy queue input parameters are known. Suppose that we are in want to determine a characteristic $\tilde{\psi}$ of a fuzzy queue whose input parameters are fuzzy numbers $\tilde{x_1}, \tilde{x_2}, ..., \tilde{x_n}$, and suppose also that ψ and $x_1, x_2, ..., x_n$ are respectively the same characteristic and the same parameters in crisp model. If the analytical ψ is known, it is often expressed by

$$\psi = f(x_1, x_2, ..., x_n) \tag{21}$$

where f is a real multivalued function using basics operations "$+, -, \times,$ and$\div$" in $\mathbb{R}$. By means of *Zadeh's* extension principle, the crisp characteristic ψ in (21) is extended to the following fuzzy characteristic

$$\tilde{\psi} = \tilde{f}(\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n) \tag{22}$$

where $\tilde{f}$ is a fuzzy multivalued function using basics fuzzy operations "$\oplus, \ominus, \otimes$ and $\oslash$ in the set of fuzzy numbers noted $\mathbb{F}(\mathbb{R})$. To determine the fuzzy queue characteristic $\tilde{\psi}$ by flexible α-cuts approach, it suffices to determine $\tilde{\psi}$ by α-cuts and intervals arithmetic. For this reason, we could use the following process:

- We determine firstly the α-cuts of all input parameters $\tilde{x_1}, \tilde{x_2}, ..., \tilde{x_n}$ under the form of closed real intervals to get $\tilde{x_1}_\alpha = [\varphi_1(\alpha), \omega_1(\alpha)], \tilde{x_2}_\alpha = [\varphi_2(\alpha), \omega_2(\alpha)], ..., \tilde{x_n}_\alpha = [\varphi_n(\alpha), \omega_n(\alpha)]$, where $\varphi_i(\alpha)$ and $\omega_i(\alpha)$ are usual real-valued functions $(1 \leq i \leq n)$.

- Applying suitable formulas from Equations (13), (14) (15), and (16) to Equation (21), we obtain

$$\tilde{\psi_\alpha} = \tilde{f}(\widetilde{x_{1\alpha}}, \widetilde{x_{2\alpha}}, ..., \widetilde{x_{n\alpha}}) \tag{23}$$

- Applying Equations (9), (10) (11), and (12) to Equation (23) we get the closed real interval

$$\tilde{\psi_\alpha} = [\tilde{\psi}^L(\alpha), \tilde{\psi}^U(\alpha)], \tag{24}$$

where $\tilde{\psi}^L(\alpha)$ and $\tilde{\psi}^U(\alpha)$ are usual real-valued functions whose reciprocals define the membership function of $\tilde{\psi}$ as follows:

$$\eta_{\tilde{\psi}}(x) = \begin{cases} \left(\widetilde{\psi}^L\right)^{-1}(x) & \text{if} \quad \widetilde{\psi}^L(0) \leq x \leq \widetilde{\psi}^L(1) \\ \left(\widetilde{\psi}^U\right)^{-1}(x) & \text{if} \quad \widetilde{\psi}^U(1) < x \leq \widetilde{\psi}^U(0) \\ 0 \quad \text{otherwise.} \end{cases} \tag{25}$$

- The real numbers $\widetilde{\psi}^L(0)$ and $\widetilde{\psi}^U(0)$ obtained in Equation (24) for $\alpha = 0$ fix the support bounds of $\widetilde{\psi}$ and indicate that $\widetilde{\psi}$ is an imprecise real number between $\widetilde{\psi}^L(0)$ and $\widetilde{\psi}^U(0)$, $\widetilde{\psi}$ does not fall below the lower bound $\widetilde{\psi}^L(0)$ or exceed the upper bound $\widetilde{\psi}^U(0)$. Its modal value $\widetilde{\psi}^L(1) = \widetilde{\psi}^U(1)$ is often taken as its most possible value.

VI. Numerical example

Let us consider a fuzzy queueing system with breakdowns and repairs whose arrivals rate, service rate, retrial rate, failure rate and repair rate are, respectively trapezoïdal fuzzy numbers $\tilde{\lambda} = (3/4/5/6)$,$\tilde{\gamma} = (25/26/27/28)$,$\tilde{\theta} = (14/15/16/17)$,$\tilde{\rho} = (36/37/38/39)$, and $\tilde{\beta} = (46/47/48/49)$. We determine the expected customers numbers in the queue, the rest of characteristics are determined similarly.
According to Equation (1), let us determine firstly the α-cuts of fuzzy parameters $\tilde{\lambda}, \tilde{\gamma}, \tilde{\theta}, \tilde{\rho}$ and , $\tilde{\beta}$ we get:

$$\tilde{\lambda}_\alpha = [\alpha + 3, -\alpha + 6], \tilde{\gamma}_\alpha = [\alpha + 25, -\alpha + 28],$$

$$\tilde{\theta}_\alpha = [\alpha + 14, -\alpha + 17], \tilde{\beta}_\alpha = [\alpha + 46, -\alpha + 49],$$

Applying α-cuts arithmetic formula to (20), we get the result: Assume that $[\tilde{N}_q]_\alpha = \frac{\tilde{E}}{\tilde{F}}$, with

$$\tilde{E} = \tilde{\lambda}_\alpha \left[\tilde{\gamma}_\alpha \tilde{\sigma}_\alpha (\tilde{\gamma}_\alpha + \tilde{\sigma}_\alpha) + \tilde{\lambda}_\alpha (\widetilde{\beta}_\alpha + \tilde{\sigma}_\alpha)^2\right]$$

and

$$\tilde{F} = \tilde{\gamma}_\alpha (\widetilde{\beta}_\alpha + \tilde{\sigma}_\alpha)[\widetilde{\beta}_\alpha (\tilde{\gamma}_\alpha + \tilde{\sigma}_\alpha) - \tilde{\lambda}_\alpha (\widetilde{\beta}_\alpha + \tilde{\sigma}_\alpha)]$$

Computation of the numerator $\tilde{E}$

$$\tilde{\gamma}_\alpha + \tilde{\sigma}_\alpha = [\alpha + 25, \alpha + 28] + [\alpha + 36, \alpha + 39] = [2\alpha + 61, 2\alpha + 67]$$

$$\begin{aligned} \widetilde{\beta}_\alpha + \tilde{\sigma}_\alpha &= [\alpha + 46, -\alpha + 49] + [\alpha + 36, -\alpha + 39] \\ &= [2\alpha + 82, 2\alpha + 88] \end{aligned}$$

$$(\tilde{\beta}_\alpha + \tilde{\sigma}_\alpha)^2 = [(2\alpha + 82)^2, (2\alpha + 88)^2]$$

$$\begin{aligned} \lambda_\alpha (\tilde{\beta}_\alpha + \tilde{\sigma}_\alpha)^2 &= [\alpha + 3, -\alpha + 6]\left[(2\alpha + 82)^2, (2\alpha + 88)^2\right] \\ &= \left[(\alpha + 3)(2\alpha + 82)^2, (-\alpha + 6)(2\alpha + 88)^2\right] \end{aligned}$$

$$\begin{aligned} \tilde{\gamma}_\alpha . \tilde{\sigma}_\alpha &= [\alpha + 25, -\alpha + 28] \,.\, [\alpha + 36, \alpha + 39] \\ &= [(\alpha + 25)(\alpha + 36), (-\alpha + 28)(-\alpha + 39)] \end{aligned}$$

$$\tilde{\gamma}_\alpha . \tilde{\sigma}_\alpha (\tilde{\gamma}_\alpha + \tilde{\sigma}_\alpha) =$$
$$[(\alpha + 25)(\alpha + 36), (-\alpha + 28)(-\alpha + 39)]\,[2\alpha + 61, -2\alpha + 67] =$$
$$[(\alpha + 25)(\alpha + 36)(2\alpha + 61), (-\alpha + 28)(-\alpha + 39)(-2\alpha + 67)]$$

Let us suppose that

$$\widetilde{K} = \tilde{\gamma}_\alpha \cdot \tilde{\sigma}_\alpha (\tilde{\gamma}_\alpha + \tilde{\sigma}_\alpha) + \tilde{\lambda}_\alpha \left(\tilde{\beta}_\alpha + \tilde{\sigma}_\alpha\right)^2$$

$$\begin{aligned} \tilde{K} =& [(\alpha + 25)(\alpha + 36)(2\alpha + 61), (-\alpha + 28)(-\alpha + 39)(-2\alpha + 67)] \\ &+ \left[(\alpha + 3)(2\alpha + 82)^2, (-\alpha + 6)(-2\alpha + 88)^2\right] \\ =& [(\alpha + 25)(\alpha + 36)(2\alpha + 61) + (\alpha + 3)(2\alpha + 82)^2, \\ & (-\alpha + 28)(-\alpha + 39).(-2\alpha + 67) + (-\alpha + 6)(-2\alpha + 88)^2] \end{aligned}$$

$$\widetilde{E} = \tilde{\lambda}_\alpha \cdot \tilde{K}$$

$$\begin{aligned} =& [\alpha+3, -\alpha+6] \cdot [(\alpha+25)(\alpha+36)(2\alpha+61)+(\alpha+3)(2\alpha+82)^2, \\ & (-\alpha + 28)(-\alpha + 39)(-2\alpha + 67) + (-\alpha + 6)(-2\alpha + 88)^2] \\ =& \left[(\alpha + 3)\left((\alpha + 25)(\alpha + 36)(2\alpha + 61) + (\alpha + 3)(2\alpha + 82)^2\right),\right. \\ & \left.(-\alpha+6)((-\alpha+28)(-\alpha+39)(-2\alpha+67)+(-\alpha+6)(-2\alpha+88)^2)\right]. \end{aligned}$$

$$\begin{aligned} \tilde{\beta}_\alpha (\tilde{\gamma}_\alpha + \tilde{\sigma}_\alpha) &= [\alpha + 46, -\alpha + 49][2\alpha + 61, -2\alpha + 67] \\ &= [(\alpha + 46)(2\alpha + 61), (-\alpha + 49)(-2\alpha + 67)], \\ \tilde{\lambda}_\alpha \left(\tilde{\beta}_\alpha + \tilde{\sigma}_\alpha\right) &= [\alpha + 3, -\alpha + 6][2\alpha + 82, -2\alpha + 88] \\ &= [(\alpha + 3)(2\alpha + 82), (-\alpha + 6)(-2\alpha + 88)], \\ \tilde{\gamma}_\alpha \left(\tilde{\beta}_\alpha + \tilde{\sigma}_\alpha\right) &= [\alpha + 25, -\alpha + 28][2\alpha + 82, -2\alpha + 88] \\ &= [(\alpha + 25)(2\alpha + 82), (-\alpha + 28)(-2\alpha + 88)]. \end{aligned}$$

Let us suppose that $\widetilde{L} = \widetilde{\beta}_\alpha (\tilde{\gamma}_\alpha + \tilde{\sigma}_\alpha) - \tilde{\lambda}_\alpha \left(\widetilde{\beta}_\alpha + \tilde{\sigma}_\alpha\right)$

$$\begin{aligned} \tilde{L} =& [(\alpha + 46)(2\alpha + 61) - (-\alpha + 6)(-2\alpha + 88), \\ & (-\alpha + 49)(-2\alpha + 67) - (\alpha + 3)(2\alpha + 82)], \end{aligned}$$

$$\widetilde{F} = \widetilde{\gamma}_\alpha \left(\widetilde{\beta}_\alpha + \widetilde{\sigma}_\alpha\right) \cdot \widetilde{L}$$

$$\begin{aligned} =& [(\alpha + 25)(2\alpha + 82), (-\alpha + 28)(-2\alpha + 88)] \cdot [(\alpha + 46)(2\alpha + 61) \\ & -(-\alpha+6)(-2\alpha+88), (-\alpha+49)(-2\alpha+67)-(\alpha+3)(2\alpha+82)] \\ =& [(\alpha + 25)(2\alpha + 82)((\alpha + 46)(2\alpha + 61) - (-\alpha + 6)(-2\alpha + 88)), \\ & (-\alpha+28)(-2\alpha+88)((-\alpha+49)(-2\alpha+67)-(\alpha+3)(2\alpha+82))] \end{aligned}$$

Using Equation (12), we can write $[\tilde{N}_q]_\alpha$ as follows:

$$\begin{aligned} \left[\tilde{N}_q\right]_\alpha = \frac{\tilde{E}}{\widetilde{F}} = \Bigg[& \min\left\{\frac{f_1(\alpha)}{g_1(\alpha)}, \frac{f_1(\alpha)}{g_2(\alpha)}, \frac{f_2(\alpha)}{g_1(\alpha)}, \frac{f_2(\alpha)}{g_2(\alpha)}\right\}, \\ & \max\left\{\frac{f_1(\alpha)}{g_1(\alpha)}, \frac{f_1(\alpha)}{g_2(\alpha)}, \frac{f_2(\alpha)}{g_1(\alpha)}, \frac{f_2(\alpha)}{g_2(\alpha)}\right\}\Bigg], \end{aligned}$$

where $\tilde{E} = [f_1(\alpha), f_2(\alpha)]$ and $\tilde{F} = [g_1(\alpha), g_2(\alpha)]$.

That is:

$$f_1(\alpha) = (\alpha+3)((\alpha+25)(\alpha+36)(2\alpha+61)+(\alpha+3)(2\alpha+82)^2)$$

$$f_2(\alpha) = (-\alpha+6)((-\alpha+28)(-\alpha+39)(-2\alpha+67)+(-\alpha+6)(-2\alpha+88)^2)$$

$$g_1(\alpha) = (\alpha+25)(2\alpha+82)((\alpha+46)(2\alpha+61)-(-\alpha+6)(-2\alpha+88))$$

$$g_2(\alpha) = (-\alpha+28)(-2\alpha+88)((-\alpha+49)(-2\alpha+67)-(\alpha+3)(2\alpha+82))$$

Finally,

$$[\tilde{N_q}]_\alpha = [\frac{f_1(\alpha)}{g_2(\alpha)}, \frac{f_2(\alpha)}{g_1(\alpha)}], \tag{26}$$

where f_1, f_2, f_3, and f_4 are the functions found above.
Basing on Equation (1), (2), and (4), $\tilde{N_q}$ is a fuzzy number whose the support and the core are respectively

$$supp(\tilde{N_q}) = [\tilde{N_q}]_0 = [\frac{f_1(0)}{g_2(0)}, \frac{f_2(0)}{g_1(0)}] =]0.03, 0.153[,$$

and

$$core(\tilde{N_q}) = [\tilde{N_q}]_1 = [\frac{f_1(1)}{g_2(1)}, \frac{f_2(1)}{g_1(1)}] = [0.0549, 0.0937].$$

These results mean that the customers number in queue belongs to the interval]0.03, 0.153[. Its most possible value belongs to the interval [0.0549, 0.0937]. According to Equation (26)30), the following membership function graphic of $\tilde{N_q}$, presented in Fig.1 below:

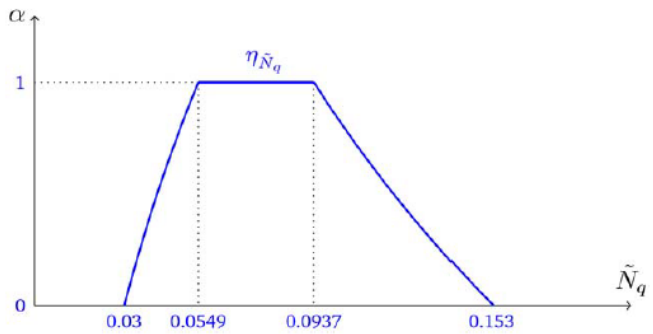


Figure 1. Membership function of expected customers number in the queue.

VII. Conclusion

This paper introduces the performance measures computation of a single server fuzzy retrial queue with breakdowns and repairs by a technique called flexible -cuts method. This one uses only one fuzzy arithmetic instead of two arithmetics used by mathematical nonlinear programming method. Basing on this approach, some characteristics as expected customers number in the queue, in the orbit or in the entire system can be computed successfully, and results are the same than those obtained by other methods.

References

[1] Aissani A. and Artalejo J. R. On the single server retrial queue subject to breakdowns,*Queueing Systems Theory Applications*, 30(1998) 309-321.
[2] R. Kalayanaraman, N. Thillaigovindan and G. Kannadasan, A single server fuzzy queue with unreliable server, International Journal of Computational Cognition, 8(1)(2010), 1-4.
[3] D. Dubois and H. Prade, Fuzzy real algebra: Some results, Fuzzy Sets and Systems 2(4) (1979), 327-348.
[4] Kulkarni V. G. and Choi Bong Dae Retrial queues with server subject to breakdowns and repairs, *Queueing Systems Theory and Applications*, 7(1990) 191-208.
[5] Z. Mueen, R. Ramli and N. Zura Zaibidi, Performance measurements of single server fuzzy queues with unreliable server using left and right method, AIP Conference Proceedings 1691(14) (2015); Article 030019.
[6] J. P. Mukeba Kanyinda. "Analysis of Retrial Fuzzy Queue with Single Server Subject to Breakdowns and Repairs by Flexible Alpha-Cuts Method." *Journal of Pure and Applied Mathematics: Advances and Applications* 22.1 (2020): 41-58.
[7] J. P. Mukeba Kanyinda, R. Mabela Makengo Matendo and B. Ulungu Ekunda Lukata, Computing fuzzy queueing performance measures by L-R method, Journal of Fuzzy Sets Valued Analysis 2015(1) (2015), 57-67.
[8] J. P. Mukeba Kanyinda, Analysis of fuzzy queue characteristics by flexible alphacuts method, Journal of Fuzzy Sets Valued Analysis 2017(1) (2017), 1-11.
[9] S. Shanmugasundaram and B. Venkatesh, An M/M/1 retrial queue with unreliable server under fuzzy environment, International Journal of Mathematical Sciences Applications 6(1) (2016), 81-88.
[10] W. Jinting, C. Jinhua and L. Quanlin, Reliability analysis of the retrial queue with server breakdowns and repairs, *Queueing Systems Theory and Applications*, 38(2001) 363-380.

Analyse du modèle M/G/1 avec rappels, clients impatients, serveur non fiable, et vacance

Ladjemil Nesrine
Unité de recherche LaMoS, Université de Béjaia
nesrine.ladjemil@univ-bejaia.dz

Rahmoune Fazia
Unité de recherche LaMoS, Université de Béjaia
fazia.rahmoune@univ-bejaia.dz

***Résumé*—**
Dans le cadre de cette étude, nous nous plongeons dans une analyse stochastique détaillée d'un modèle d'attente M/G/1 avec rappels, clients impatients, un serveur sujet à des pannes, et des périodes de vacance. Notre approche méthodologique repose sur l'utilisation de chaînes de Markov pour élaborer l'expression de la distribution stationnaire qui caractérise l'état du système. Cette distribution stationnaire revêt une importance cruciale en nous permettant de calculer une série de mesures de performance du système. Ainsi, notre recherche se base sur une méthodologie analytique rigoureuse pour explorer en profondeur les complexités de ce modèle d'attente dans divers contextes opérationnels.

***Mots-Clés*—**
Serveur non fiable, rappels, impatience, maintenance, vacance, mesure de performances.

I. Introduction

La modélisation est devenue une pratique de plus en plus courante pour concevoir et analyser divers systèmes réels. La théorie des files d'attente est une méthode fondamentale dans la modélisation stochastique, visant à évaluer les performances et à contrôler des systèmes variés tels que la production industrielle et les systèmes informatiques. [1], [2], [3]
On peut représenter un système ou un phénomène d'attente de la manière suivante : un groupe d'individus, appelés clients, arrive de manière aléatoire ou suivant un certain processus pour recevoir un service d'un autre individu désigné comme serveur. La formation de la file d'attente débute dès que le taux d'arrivée des clients dépasse le taux de service (où par "taux", on entend le nombre moyen de clients arrivant ou étant servis par unité de temps). La file d'attente peut également se former de manière abstraite, comme une file de machines en panne attendant d'être réparées dans un atelier, ou un ensemble de programmes attendant l'acquisition d'un composant de la machine, etc. Ainsi, un système de files d'attente se compose d'un espace de service avec un ou plusieurs dispositifs de service (serveurs), ainsi que d'un espace d'attente où une file d'attente éventuelle se forme.

Il existe une vaste littérature consacrée aux systèmes de files d'attente [4], [5], [6]. Les premiers travaux sur ce sujet sont attribués au mathématicien danois Erlang (1878-1929). Ils ont débuté lorsqu'il s'est intéressé à modéliser le réseau téléphonique de Copenhague pour réduire les temps d'attente. Par la suite, la théorie mathématique des files d'attente a connu un développement significatif grâce aux contributions de Palm, Kolmogorov et Khintchine.

La théorie classique des files d'attente propose deux solutions pour résoudre le conflit qui survient lorsque qu'un client arrive dans un système à un seul serveur et trouve ce dernier occupé : soit le client quitte le système sans recevoir le service, soit il intègre une file d'attente. Une alternative consiste à autoriser le client à renouveler sa demande de service après un laps de temps aléatoire. Entre deux tentatives successives (rappels), le client reste en attente. Un tel système est appelé un système de files d'attente avec rappels.

Cette branche particulière de la théorie des files d'attente est d'une grande importance dans la modélisation stochastique de divers problèmes rencontrés dans les domaines des télécommunications et des réseaux informatiques. Les systèmes de files d'attente avec rappels se révèlent être des outils essentiels pour aborder des questions pratiques variées. Ils permettent par exemple d'analyser le comportement des abonnés au sein des réseaux téléphoniques, de mettre en place des mécanismes d'évitement de collisions dans les réseaux locaux, d'étudier les temps d'attente pour accéder à la mémoire des disques magnétiques, et bien d'autres applications encore [15]. En conséquence, un grand nombre de recherches ont été menées et publiées dans des revues spécialisées couvrant des domaines tels que les probabilités appliquées, les modèles stochastiques, les statistiques, la recherche opérationnelle, les télécommunications, l'ingénierie industrielle et l'informatique.

L'engouement pour ce domaine est si prononcé qu'il s'est traduit par l'organisation de plusieurs ateliers dédiés aux systèmes de files d'attente avec rappels, tenus dans des villes telles que Madrid (1998), Minsk (1999), Amsterdam (2000), Cochin (2002), Seoul (2004),

Miraflores de la Sierra (2006), Athens (2008) et Beijing (2010). En reconnaissance de cette importance, certaines revues de renommée internationale ont consacré des numéros spéciaux à ce sujet, notamment le journal "Annals of Operation Research" [17] et "Mathematical and Computer Modelling" [16].
Les modèles d'attente avec rappels ont fait l'objet de nombreuses études. Parmis les premières contributions sur le sujet, on trouve celle de Cohen [12] publiée en 1957 dans "Philips Telecommunication Review", de Elldin [11] en 1967 dans "Erricsson Technics" et de Hashida et Kawashima [10] dans "Electronics and Communication in Japan". Les progrès récents sont résumés dans les articles de synthèse de Aïssani(1994) [13], Kulkani et Liang (1997) [20], Templeton (1999) [21] et dans les monographies de Falin et Templeton (1997) [18], Gòmez-Corral et Ramalhoto (2000) [19] et dans les travaux bibliographiques de Artalejo (1999 et 2010) [16], [14].
Lorsqu'on explore les problèmes classiques de la théorie des files d'attente, l'hypothèse courante était que les serveurs étaient entièrement fiables. Cependant, dans la réalité, on se heurte souvent à des situations où les serveurs peuvent subir des pannes de manière aléatoire, interrompant ainsi le service aux clients pendant un certain laps de temps. L'étude de tels systèmes est indubitablement cruciale pour des applications concrètes.

Dans le domaine de la théorie des files d'attente, nous nous penchons également sur l'analyse de l'impact de la non-fiabilité des serveurs sur les caractéristiques du système en question. Ainsi, les modèles mathématiques les plus sophistiqués des systèmes et réseaux de files d'attente sont justement ceux qui prennent en considération la possibilité de pannes des serveurs.

Plusieurs chercheurs ont étudié les systèmes de files d'attente où les serveurs peuvent subir des pannes et nécessiter des réparations, parmi eux Gaver [9], Avi-Itzhak and Naor [7], et d'autres. On peut trouver une bibliographie exhaustive sur le sujet dans la synthèse réalisée par Fiems et al. [8].

Dans le but d'effectuer une analyse mathématique d'un système de file d'attente, il est essentiel d'introduire un processus stochastique qui décrit l'état du système à un instant donné. En général, il existe deux catégories de processus stochastiques décrivant l'état d'un système de file d'attente : les processus stochastiques markoviens et les processus non-markoviens. Toutefois, plusieurs méthodes permettent de transformer ces derniers en processus markoviens grâce à certaines transformations (comme la méthode de la chaîne de Markov induite ou la méthode des variables supplémentaires).

En fonction des grandeurs qui déterminent la structure du système, l'objectif est de calculer le régime transitoire. Cependant, il est souvent observé que le calcul explicite de ce régime est difficile voire impossible pour la plupart des modèles. Par conséquent, il est souvent plus judicieux de se concentrer sur la détermination du régime stationnaire. La distribution stationnaire du processus stochastique introduit permet d'obtenir les indicateurs de performance du système tels que le temps d'attente d'un client, le nombre moyen de clients dans le système, le taux d'occupation des dispositifs de service, et bien d'autres encore [6].

Il existe plusieurs versions des files d'attente avec rappels, incluant des clients impatients et des interruptions de service. Les clients peuvent quitter après une attente aléatoire. En pratique, les serveurs connaissent des interruptions, ce qui peut être dû à diverses raisons telles que des pannes, des vacances du serveur, l'arrivée de clients prioritaires ou des interruptions induites par le client. Depuis les années 1980, de nombreuses études ont été menées sur les files d'attente avec rappels non fiables, abordant des pannes actives ou passives pendant le fonctionnement ou l'inactivité du serveur. La maintenance préventive est essentielle pour améliorer les performances et réduire les interruptions, mais son application spécifique dans les files d'attente avec rappels n'a pas été pleinement explorée. Ce travail se focalise sur une nouvelle version innovante de ces files d'attente intégrant la maintenance corrective et préventive, en considérant des clients persistants et impatients. Dans cette étude, nous examinons un système de files d'attente M/G/1 avec rappels, clients impatients, un serveur sujet à des pannes actives et passives, et des périodes de vacance. Nous considérons deux types d'appels primaires : persistants et impatients. Deux types de maintenance sont implémentés : la maintenance préventive, représentée par des périodes de vacance, est programmée régulièrement pour améliorer les performances du système. La maintenance corrective (ou réparation) est déclenchée lorsque le serveur tombe en panne. Notre objectif initial est d'évaluer la condition d'ergodicité de ce modèle, une étape cruciale avant d'analyser en profondeur ses performances.

Ce modèle complexe, lié à la modélisation des systèmes de fiabilité, présente un intérêt particulier pour la gestion des systèmes d'attente dans divers domaines. Les résultats de cette recherche pourraient fournir des éclaircissements importants pour optimiser les politiques de maintenance, améliorer la performance et la fiabilité des systèmes d'attente, et ainsi contribuer à des bénéfices économiques et opérationnels considérables.

II. Description du modèle M/G/1 avec rappels, clients impatients, serveur non fiable et vacance

Explorons un système de files d'attente de type M/G/1 avec rappels, comportant deux catégories d'appels principaux : persistants et impatients. Le serveur est sujet à des pannes actives et passives, ainsi qu'à des périodes de vacance. Voici les caractéristiques du système :

- Les arrivées primaires sont distribuées selon un processus de poisson telles que :
  - Les arrivées persistantes suivent la loi de poisson de paramètre λ.
  - Les arrivées impatientes suivent la loi de poisson de paramètre γ.
- Les temps de service des clients persistants sont indépendants avec la fonction de repartition $H(x)$, sa transformée de laplace stieltjes $h(s)$ et les premiers moments donnés par $h_1 h_2$.
- Les temps de service des clients impatients sont indépendants avec la fonction de repartition $F(x)$, sa transformée de laplace stieltjes $f(s)$ et les premiers moments donnés par $f_1 f_2$.
- Les rappels (arrivées secondaires) : chaque client en orbit revient indépendamment des autres après une durée exponnentielle avec le paramètre τ.
- Le serveur sujet à des pannes passives ou actives, elles se produisent selon un processus de poisson avec les taux θ_0 et θ_1.
- Deux formes de maintenance sont mises en œuvre : la maintenance préventive et la maintenance corrective. :
  - **Maintenance corrective (réparation retardée) :** Une fois qu'une panne se réalise, le serveur devient inactif et attend un temps aléatoire pour activer le processus de réparation, Ce temps d'attente est exponentiellement distribué avec un taux κ. La durée de réparation est exponentiellement distribuée avec un taux r. le serveur tombe en panne soit, pendant la période d'activité ou d'inactivité.
    1) Pendant la période d'activité : sa durée est une variable aléatoire avec une fonction de distribution de probabilité $R_1(x)$ et transformée de laplace stieltjes $r_1(s)$ et les premiers moments r_{11} et r_{12}.
    2) Pendant la période d'inactivité : sa durée est une variable aléatoire avec une fonction de distribution de probabilité $R_0(x)$ et TL-S $r_0(s)$ et les premiers moments r_{01} et r_{02}.
  - **maintenance préventive :** les durées de cette maintenance sont similaires à celles des periodes de vacance du serveur. Ainsi dans ce cas :
    •Le serveur a la possibilité de prendre une vacance une fois que tous les clients ont été servis, c'est-à-dire lorsque la file est vide. Dans ce cas, on dit que lors du début de la période de vacance, le serveur applique la politique de service exhaustif.
    • Si, à la fin de sa période de vacance, le serveur constate que la file est vide, il reprend sa vacance. En revanche, s'il constate que la file est occupée, il passe immédiatement en période d'activité, marquant ainsi la fin de sa vacance et utilisant la vacation multiple..
    • **Durée de vacance :** on note $V(x)$ la fonction de repartition des durées de vacance avec un taux ν, sa transformée de laplace stieltjes $V(x)$.

Toutes les variables aléatoires sont supposées mutuellement indépendantes et tout les moments sont supposés finis. L'état du système est,décrit par le processus :

$$\psi(t) = \{\alpha(t), \beta(t), \varepsilon(t), N(t)\}, t \geq 0,$$

où, $N(t)$ est le nombre de clients en orbit à l'instant t.

$$\alpha(t) = \begin{cases} 0, & \text{si le serveur est libre à la date } t; \\ 1, & \text{si le serveur est occupé par un client persistant;} \\ 2, & \text{si le serveur est occupé par un client impatient.} \end{cases}$$

$$\beta(t) = \begin{cases} 0, & \text{si le serveur est fonctionnel à la date } t; \\ 1, & \text{si le serveur est en panne et en attente du} \\ & \text{commencement du processus de réparation;} \\ 2, & \text{si le serveur est en état de réparation;} \\ 3, & \text{si le serveur est en vacances.} \end{cases}$$

$$\varepsilon(t) = \begin{cases} 0, & \text{si le serveur est} \\ & \text{libre à la date } t; \\ \text{temps de service écoulé,} & \text{si } \beta(t) = 0 \text{ et} \\ & \alpha(t) \neq 0; \\ \text{temps de maintenance écoulé,} & \text{si } \beta(t) = 1 \text{ ou} \\ & \beta(t) = 2; \\ \text{temps de vacance écoulé,} & \text{si } \beta(t) = 3. \end{cases}$$

Soit t_n l'instant durant lequel le serveur devient inactif pour la n^{eme} fois, donc on observe le processus $\zeta(t)$ au instant t_n, qui correspondent aux instants de fin de service. On a donc $q_n = N(t_n)$, le nombre de clients persistants dans l'orbit juste après cet instant.
• Soit une suite d'intervalles (cycles) successifs $]t_{n-1}, t_n[$, $n \geq 1$. de différents types.

- **Type 01 :** Le serveur subit une panne passive avant une arrivée.

— **Type 02 :** Le serveur prend une vacance avant une panne ou une arrivée primaire.

— **Type 03 :** Un client persistant est pris en charge jusqu'à ce qu'une panne, une maintenance ou un rappel survienne. Lorsqu'une panne active se produit, le temps de service est interrompu, ce qui entraîne un retard dans la mise en œuvre de la réparation requise.

— **Type 04 :** un client persistant à lieu avant une panne ou une maintenance ou un rappel, le temps de service est accompli.

— **Type 05 :**Lorsqu'un client impatient se présente et qu'une panne, une maintenance ou un rappel survient simultanément, le temps de service est interrompu en raison de la panne active, ce qui engendre un retard dans la mise en œuvre de la réparation requise.

— **Type 06 :** un client impatient à lieu avant une panne ou une maintenance ou un rappel, le temps de service est accompli.

— **Type 07 :** Lorsqu'un rappel intervient en premier parmi les événements, le temps de service subit une interruption à cause d'une panne active, entraînant un délai dans la mise en œuvre de la réparation nécessaire.

— **Type 08 :** Un rappel se produit avant tous les autres évenements, le temps de service est accompli.

Soit $I_n = k$, le n^{eme} cycle est de type k, La chaîne de Markov induite $(q_n)_{(n\geq 1)}$ satisfait l'équation récursive suivante :

$$q_{n+1} = \begin{cases} q_n + A_I^{(n+1)} + \\ A_R^{(n+1)} + A_{AR}^{(n+1)}, & \text{si } I_{n+1} = 1, \\ q_n + A_V^{(n+1)} + A_1^{(n+1)}, & \text{si } I_{n+1} = 2, \\ q_n + A_1^{(n+1)} + A_P^{(n+1)} + A_R^{(n+1)} + \\ A_{AR}^{(n+1)} + 1, & \text{si } I_{n+1} = 3, \\ q_n + A_1^{(n+1)}, & \text{si } I_{n+1} = 4, \\ q_n + A_P^{(n+1)} + A_2^{(n+1)} + \\ A_R^{(n+1)} + A_{AR}^{(n+1)}, & \text{si } I_{n+1} = 5, \\ q_n + A_2^{(n+1)}, & \text{si } I_{n+1} = 6, \\ q_n + A_1^{(n+1)} + A_P^{(n+1)} + \\ A_R^{(n+1)} + A_{AR}^{(n+1)}, & \text{si } I_{n+1} = 7, \\ q_n + A_1^{(n+1)} - 1, & \text{si } I_{n+1} = 8. \end{cases}$$

Dans un contexte général, l'expression de la chaîne de Markov induite se présente de la manière suivante :

$$q_{n+1} = \begin{cases} q_n + A^{(n+1)} - B_i, & \text{si } q_n > 0 \text{ et } I_{n+1} \neq 3, \\ A_V^{(n+1)}, & \text{si } q_n = 0, \\ q_n + A^{(n+1)} + 1, & \text{si } q_n > 0 \text{ et } I_{n+1} = 3. \end{cases}$$

Telsque,

— $A_{(n+1)}$, le nombre de clients persistants qui entrent juste après l'instant t_n (le départ de n^{eme} client).

— q_n, le nombre de clients peristants dans l'orbit justé après l'instant t_n.

— $A_I^{(n)}$ le nombre d'arrivées persistantes durant la panne passive dans le n^{eme} cycle de type 1.

— $A_V^{(n)}$ le nombre d'arrivées persistantes durant la période de vacance dans le n^{eme} cycle de type 2.

— $A_P^{(n)}$ le nombre d'arrivées persistantes durant la panne active dans le n^{eme} cycle de types 3, 5 et 7.

— $A_1^{(n)}$ le nombre d'arrivées persistantes durant le temps de service d'un client persistant jusqu'à qu'il soit interrompu par une panne active dans le n^{eme} cycle de type 4.

— $A_2^{(n)}$ le nombre d'arrivées persistantes durant le temps de service d'un client impatient jusqu'à qu'il soit interrompu par une panne active dans le n^{eme} cycle de type 6.

— $A_R^{(n)}$ le nombre d'arrivées persistantes durant période de réparation dans le n^{eme} cycle de type 1, 3, 5 et 7.

— $A_{AR}^{(n)}$ le nombre d'arrivées persistantes juste après la réparation dans le n^{eme} cycle de type 1, 3 ,5 et 7.

— Et,

$$B_i = \begin{cases} 0, & \text{si le client qui quitte le système à} \\ & \text{l'instant } t_{n+1}\text{provient de l'exterieur,} \\ 1, & \text{si le client qui quitte le système à} \\ & \text{l'instant } t_{n+1}\text{provient de l'orbit,} \end{cases}$$

III. Conclusion

Cette étude ambitieuse vise à approfondir notre compréhension d'un système d'attente complexe, en tenant compte des éléments tels que les rappels, le serveur non fiable, les clients impatients et les périodes de vacances. Cependant, il est important de noter que cette recherche est encore en cours, et notre démarche se divise en plusieurs étapes essentielles.

La première étape de cette étude consistera à évaluer la condition d'ergodicité du modèle d'attente. Cette condition est fondamentale pour s'assurer que le système atteint un état d'équilibre stable, ce qui est nécessaire pour des analyses plus approfondies des performances. La vérification de cette condition est cruciale pour garantir la validité de notre modèle et l'obtention de résultats significatifs. Une fois cette condition évaluée, nous serons en mesure de passer à l'étape suivante, qui consiste à explorer en détail les performances du modèle.

RÉFÉRENCES

[1] W. BUX, " Local area subnetworks : a performance comparison". *IEEE, Transaction on communications*, 1981, 29(10), 1465-1473.

[2] A. AISSANI And J.R. ARTALEJO. "On the single server retrial queue subject to breakdowns", *Queueing Systems*, 1998, 30(3-4), 309-321.

[3] S.K.S. GUPTA and P.K. SRIMANI, "Scheduling independent jobs for torus connected network with/ without link contention". *Mathematical and computer modeling*, 2000, 131-140.

[4] G. I. FALIN, "Single-line repeated orders queueing systems", *Mathematische Operations Forchung und Statistik, Optimization*, 1986, pp. 649-667.

[5] P. P. BOCHAROV, O. I. PAVLOVA and D. A. PUZIKOVA, "M/G/1/r retrial queueing systems with priority of primary customers," *Mathematical and Computer Modelling*, 1999 , 30, pp. 89-98.

[6] A. RUEGG, " Processus Stochastiques". *Presses polytechniques romandes*, Lausannes, 1989.

[7] B. AVI-ITZHAK and P. NAOR. "Some Queueing Problems with the service station subject to breakdown." *Oper.Res.*,1963, 11(3) :303–320.

[8] D. FIEMS, B. STEYAERT, and H. BRUNEEL. "Discrete-time queues with generally distributed service times and renewal-type server interruptions." *Performance Evaluation*, 2004, 55 :277–298.

[9] D. P. GAVER. "A waiting Line with Interrupted Service, including Priorities". *Roy. Stat. Soc. J*, 1962, B25 :73–90.

[10] O. HASHIDA and K. KAWASHIMA. "Buffer Behavior with Repeated Calls".*Electronics and Communication in Japan*, 1979, 62-B, 27.

[11] A. ELLDIN. "Approach to the theoretrial description of repeated call attempts". *Erricsson Technics*, 1967, 23(3), 346-407.

[12] J.W. COHEN, " Basic problems of the telephone traffic theory and the influence of repeated calls". *Philis Telecom, Review*, 1957, 18(2), 49-100.

[13] A. AISSANI, "Survey on Retrial Queueing Models".*Actes des Journées Statistiques Appliquées*, U.S.T.H.B., Alger, 1994, 1-11.

[14] J.R. ARTALEJO, "Accessible bibliography on retrial queues : Progress in 2000-2009". *Mathematical and Computer Modelling*, 2010, 51, 1071-1081.

[15] J.R. ARTALEJO, "Retrial queues with a finite number of sources". *Journal of the Korean Mathematical Society*,1998, 35(3) :503-525.

[16] J.R. ARTALEJO, "Retrial Queueing systems", *Mathematical and Computer Modelling 30*, 1999, No. 3-4, 1-228.

[17] J.R. ARTALEJO, "Algorithmic Methods in Retrial Queues". *Annals of Operation Research*, 2006, 141,1-301

[18] G.I. FALIN and J.G. TEMPLETON, "Retrial queues". *New Jersey : Chapman and Hill*, 1997.

[19] A. GOMEZ-CORRAL and M.F. RAMALHOTO, "On the waiting time distribution and the busy period of a retrial queue with constant retrial rate". *Stochastic Modelling and Applications*, 2000, 3, 37-47.

[20] V.G. KULKARNI and H.M. LIANG, "Retrial Queues Revisited". *Frontiers in Queueing (J.H. Dshalov, ed.) CRC Press Boca Raton*, 1997, pp 19-34.

[21] James. G. C. TEMPLETON, "Retrial Queues", *Top, vol*, 1999, 7, p. 351-353.

V

ESTIMATION NON PARAMÉTRIQUE DE LA FIABILITÉ

Sommaire

Nonparametric availability density function estimation using gamma kernel

Zitout Yasmina
Laboratory LMA, University of Bejaia, Bejaia, Algeria
yasmina.zitout@univ-bejaia.dz

Lagha Karima
Research Unit LaMOS, University of Bejaia, Bejaia, Algeria
karima.lagha@univ-bejaia.dz

***Abstract*—**
In this paper, we propose a nonparametric kernel estimator for availability density function of a repairable system based on gamma kernel in the context of positively skewed data. Some properties: bias, variance, MSE (Mean Squared Error) and MISE (Mean Integrated Squared Error) of the proposed estimator are also investigated. In addition two popular approaches given by unbiased cross validation and rule of thumb are adapted for bandwidth selection. Finally, a simulation study is performed to assess accuracy of the estimator.

***Keywords*—**
Reliability, Gamma kernel, Bandwidth parameter, Availability.

I. Introduction

The last two decades have seen major advances in the developments of new maintenance strategies. The main goals of these strategies are to reduce equipment downtime as well as increase equipment availability. Availability is the most common measure of the effectiveness of a repairable (maintained) system, it measures the probability that a system will not fail or undergo a repair action when it must be used, there are several types, namely : instantaneous, average, asymptotic, etc. It takes its place in various fields : energy, mechanics, medicine, etc.

Availability estimation has been widely discussed in recent years by many authors in both approaches (parametric and nonparametric). The parametric estimation task is carried out using standard estimation, we can cite Jacobson and Arrora [1], Gaver and Chu [2], Vasquez et al [3] and among the works of nonparametric estimation, we cite Baxter and Li [4], Huang and Mi [5]. The main objective is to propose an estimator of asymptotic availability density function using Gamma kernel considering a repairable system.

II. Gamma kernel availability density estimator

Consider a repairable component with unknown uptime, represented by a random variable (r.v) Z of unknown pdf f_Z and repair time with mean α, $\alpha > 0$ (supposed known).

The asymptotic availability is a random variable A, defined as follows

$$A = \frac{Z}{Z+\alpha} \tag{1}$$

The unknown density function f_A is given by

$$f_A(z) = \frac{(\alpha+z)^2}{\alpha} f_Z(z), \qquad z > 0, \qquad \alpha > 0. \tag{2}$$

The kernel estimator of the pdf f_A is given as follows

$$\hat{f}_{A,K(z,h)}(z) = \frac{(\alpha+z)^2}{\alpha} \hat{f}_{Z,K(z,h)}(z), \qquad z > 0, \qquad \alpha > 0 \tag{3}$$

Where $K(z,h)$ represent a Gamma kernel function (Chen [6]) and h a bandwidth parameter such as $h = h(n) > 0$ satisfying $\lim_{n\to\infty} h = 0$ and $\lim_{n\to\infty} nh = \infty$.

The gamma kernel availability density function estimator is given as follows

$$\hat{f}_{A,K_{G(z,h)}}(z) = \frac{(\alpha+z)^2}{n\alpha\Gamma(z+1)h^{\frac{z}{h}+1}} \sum_{i=1}^{n} Z_i^{\frac{z}{h}} e^{-\frac{Z_i}{h}}, \quad z > 0 \tag{4}$$

where $\Gamma(.)$ is the gamma function defined as $\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt, \quad x > 0$

A. Properties of the estimator

Under the following conditions we establish the properties of the proposed kernel estimator defined in eq. 4

C1. f is twice differentiable and its second derivative is continuous and bounded;

C2. $h = h(n)$ satisfying $h \to 0$ and $nh \to \infty$ as $n \to \infty$;

C3. $\int_0^{+\infty} z^4 \left[f'(z) + \frac{1}{2}zf''(z)\right]^2 dz < \infty$ and $\int_0^{+\infty} z^{\frac{7}{2}} f(z)dz < \infty$.

Theorem 1. *(Bias, variance and MSE of $\hat{f}_{A,K_{G(z,h)}}$)*
Under condition C_1, the bias, variance and MSE of $\hat{f}_{A,K_{G(z,h)}}$ are given by

$$Bias(\hat{f}_{A,K_{G(z,h)}}(z)) = \frac{(\alpha+z)^2}{\alpha} h\left[f'(z) + \frac{1}{2}zf''(z)\right] + o(h), \tag{5}$$

$$\mathbb{V}ar(\hat{f}_{A,K_G(z,h)}(z)) = \frac{(\alpha+z)^4}{2\alpha^2\sqrt{\pi}} n^{-1}h^{-\frac{1}{2}}z^{-\frac{1}{2}}f(z) + o(n^{-1}h^{-\frac{1}{2}}) \tag{6}$$

$$MSE\ (\hat{f}_{A,K_G(z,h)}(z)) = \left(\frac{h}{\alpha}\right)^2 (\alpha+z)^4 \left[f'(z) + \frac{1}{2}zf''(z)\right]^2 + \frac{(\alpha+z)^4}{2\alpha^2\sqrt{\pi}} n^{-1}h^{-\frac{1}{2}}z^{-\frac{1}{2}}f(z) + o(h^2 + n^{-1}h^{-\frac{1}{2}}) \tag{7}$$

Theorem 2. *(MISE of $\hat{f}_{A,K_{G(z,h)}}$)*

Under conditions C_1, C_2 and C_3 we obtain the MISE of $\hat{f}_{A,K_{G(z,h)}}$ as follows

$$MISE(\hat{f}_{A,K_{G(z,h)}}) = \left(\frac{h}{\alpha}\right)^2 \int_0^{+\infty} (\alpha+z)^4 \left[f'(z) + \frac{1}{2}zf''(z)\right]^2 dz + \frac{n^{-1}h^{-\frac{1}{2}}}{2\alpha^2\sqrt{\pi}} \int_0^{+\infty} (\alpha+z)^4 z^{-\frac{1}{2}} f(z)\ dz + o(h^2 + n^{-1}h^{-\frac{1}{2}}) \tag{8}$$

By minimizing eq. 8 in the bandwidth h, we obtain the optimal value

$$h_G^{opt} = \left[\frac{\frac{1}{4\sqrt{\pi}} \int_0^{+\infty} (\alpha+z)^4 z^{-\frac{1}{2}} f(z)\ dz}{2\int_0^{+\infty} (\alpha+z)^4 \left[f'(z) + \frac{1}{2}zf''(z)\right]^2 dz}\right]^{\frac{2}{5}} n^{-\frac{2}{5}} \tag{9}$$

Substituting the optimal bandwidth, we get the optimal mean integrated squared error as follows

$$MISE^*(\hat{f}_{A,K_{G(z,h)}}) = \frac{5}{2^{\frac{8}{5}}\alpha^2} \left[\frac{1}{2\sqrt{\pi}} \int_0^{+\infty} (\alpha+z)^4 z^{-\frac{1}{2}} f(z) dz\right]^{\frac{4}{5}} \times \left[\int_0^{+\infty} (\alpha+z)^4 \left[f'(z) + \frac{1}{2}zf''(z)\right]^2 dz\right]^{\frac{1}{5}} n^{-\frac{4}{5}} \tag{10}$$

III. Convergence properties of gamma kernel estimator

In this section, we present the weak consistency in L_1, uniform weak consistency, uniform almost sure consistency and asymptotic normality of availability density function estimator given in the following propositions. We assume that

Cond1. The second derivative of f is continuous and bounded on $[0, +\infty[$;

Cond2. $h \to 0$, $n\sqrt{h} \to \infty$ as $n \to \infty$;

Cond3. $\lim_{n\to+\infty} h = 0$ and $\lim_{n\to+\infty} nh^2 = +\infty$.

Proposition 1. *(Weak consistency in L_1 of $\hat{f}_{A,K_{G(z,h)}}$) Under Cond3*

$$\int_0^{+\infty} |\hat{f}_{A,K_{G(z,h)}}(z) - f(z)|\ dz \xrightarrow{P} 0 \quad as \quad n \to \infty \tag{11}$$

Proposition 2. *(Uniform weak consistency of $\hat{f}_{A,K_{G(z,h)}}$) Let I a compact set in $[0, +\infty[$, under cond3*

$$\underset{z\in I}{Sup} |\hat{f}_{A,K_{G(z,h)}}(z) - f(z)| \xrightarrow{P} 0 \quad as \quad n \to \infty \tag{12}$$

Proposition 3. *(Uniform almost sure consistency of $\hat{f}_{A,K_{G(z,h)}}$)*

Under Cond1 and Cond2, we assume that $\left(\frac{logn}{n\sqrt{h}}\right) \to 0$, then for any constants a and b such that $0 < a < b < \infty$ the uniform almost sure consistency of $\hat{f}_{A,K_{G(z,h)}}$ is given by

$$\underset{z\in[a,b]}{Sup} |\hat{f}_{A,K_{G(z,h)}}(z) - f(z)| = O(h) + o\left(\frac{\sqrt{logn}}{\sqrt{n\sqrt{h}}}\right) \quad a.s \tag{13}$$

Proposition 4. *(Asymptotic normality of $\hat{f}_{A,K_{G(z,h)}}$) Under Cond1 and Cond2, the estimator $\hat{f}_{A,K_{G(z,h)}}$ converges to the normal distribution as follows*

$$\left(\frac{(\alpha+z)^4 f_Z(z)}{2n\alpha^2\sqrt{\pi z h}}\right)^{-\frac{1}{2}} \left[\hat{f}_{A,K_{G(z,h)}}(z) - f_{A,K_{G(z,h)}}(z)\right] - \left(\frac{(\alpha+z)^4 f_Z(z)}{2n\alpha^2\sqrt{\pi z h}}\right)^{-\frac{1}{2}} \left[\frac{(\alpha+z)^2}{\alpha} h\left(f'_Z(z) + \frac{1}{2}zf''_Z(z) + o(h)\right)\right] \tag{14}$$

B. Bandwidth selection

As the optimal bandwidth defined in eq. 9 depends on the unknown quantities f, f' and f'' which cannot be calculated in practice, we adopt two popular methods which make it possible to solve this problem such as RT and UCV.

1) Rule of Thumb (RT): The optimal bandwidth given by a gamma reference model is given by

$$h_G^{RT} = \left[\frac{\frac{1}{4\sqrt{\pi}} \int_0^{+\infty} (\alpha+z)^4 z^{-\frac{1}{2}} f_{G(\hat{a},\hat{b})}(z)\ dz}{2\int_0^{+\infty} (\alpha+z)^4 \left[f'_{G(\hat{a},\hat{b})}(z) + \frac{1}{2}zf''_{G(\hat{a},\hat{b})}(z)\right]^2 dz}\right]^{\frac{2}{5}} n^{-\frac{2}{5}} \tag{15}$$

Other reference models can be used such as lognormal, weibull, GBS, etc.

Table I: Some expected values of ISE for availability density function estimator, based on 100 replications and different values of α.

		Gamma (8,1)		**Weibull (4,10)**		**Lognormal (2,0.4)**		**Gompertz (0.005,0.3)**	
n	α	h_{ucv}	h_{RT}	h_{ucv}	h_{RT}	h_{ucv}	h_{RT}	h_{ucv}	h_{RT}
50	0.03	0.140933	0.123418	0.051471	0.070827	0.179024	0.218658	0.041279	0.035566
	0.5	**0.000490**	**0.003482**	**0.000203**	**0.001071**	**0.000748**	**0.003347**	**0.000592**	**0.000676**
	2	**0.000580**	**0.001584**	**0.000185**	**0.000458**	**0.000831**	**0.002031**	**0.001412**	**0.000859**
	50	0.133932	0.156218	0.061252	0.054778	0.175434	0.184574	0.374274	0.333888
100	0.03	0.066402	0.045950	0.021433	0.019218	0.084845	0.080165	0.027270	0.029402
	0.5	**0.000436**	**0.001262**	**0.000159**	**0.000296**	**0.000617**	**0.001299**	**0.000468**	**0.000441**
	2	**0.000429**	**0.001186**	**0.000155**	**0.000429**	**0.000576**	**0.001618**	**0.000700**	**0.000752**
	50	0.091350	0.058052	0.033155	0.027902	0.121886	0.092427	0.221326	0.236408
500	0.03	0.033685	0.010581	0.010369	0.004257	0.043443	0.013814	0.021766	0.042565
	0.5	**0.000370**	**0.000216**	**0.000117**	**8.386464e^{-5}**	**0.000478**	**0.000238**	**0.000367**	**0.000654**
	2	**0.000484**	**0.000299**	**0.000144**	**9.153435e^{-5}**	**0.000549**	**0.000313**	**0.000567**	**0.000856**
	50	0.075973	0.070090	0.025566	0.157841	0.090922	0.104817	0.184702	5.639185
1000	0.03	0.028105	0.010916	0.008974	0.005783	0.041368	0.013629	0.021901	0.080641
	0.5	**0.000393**	**0.000204**	**0.000124**	**0.000103**	**0.000521**	**0.000227**	**0.000320**	**0.001447**
	2	**0.000428**	**0.000275**	**0.000132**	**0.000136**	**0.000544**	**0.000246**	**0.000597**	**0.002721**
	50	0.068042	1.062695	0.022377	0.587378	0.086813	0.296817	0.172368	23.51793

2) Unbiased Cross Validation (UCV): The optimal bandwidth established by the unbiased cross validation method is given as follows

$$h_G^{ucv} = \underset{h>0}{argmin} \int_0^{+\infty} \left[\frac{(\alpha+z)^2}{\alpha n} \sum_{i=1}^{n} K_G(Z_i) \right]^2 dz - \frac{2(\alpha+z)^2}{\alpha n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{G(Z_i,h)}(Z_j) \quad (16)$$

where $K_G(Z_i)$ is the gamma kernel.

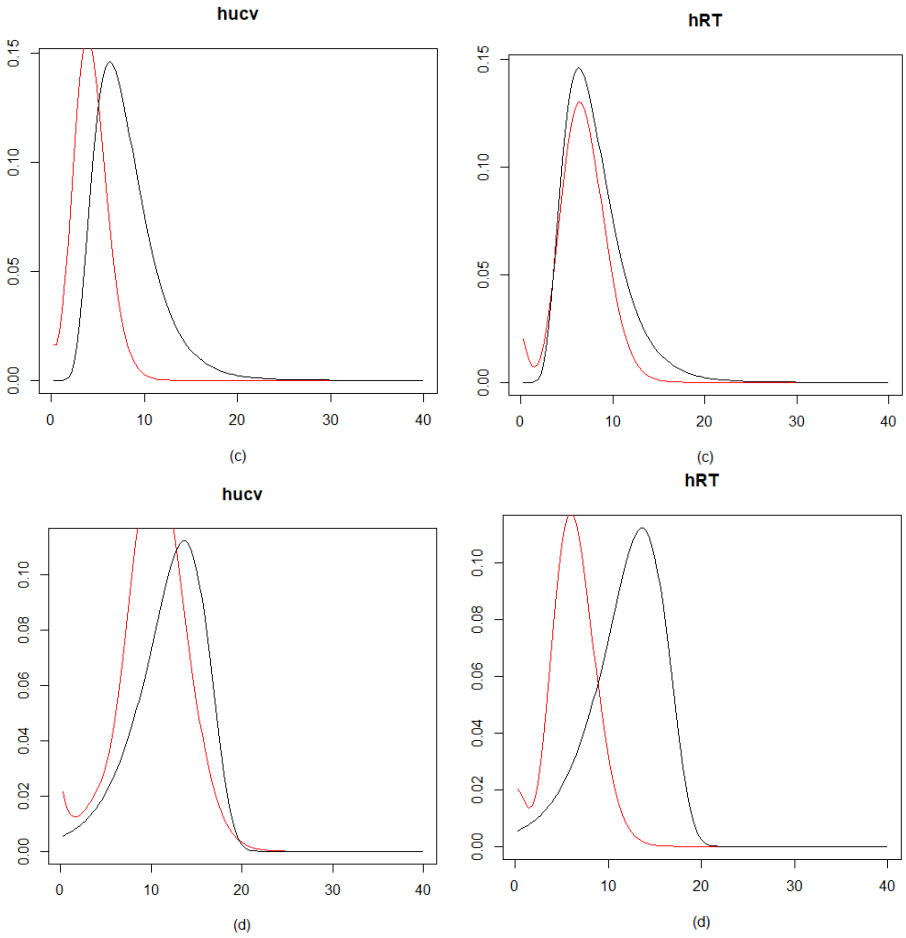


Figure 1: Plots of gamma (red) availability density function estimator for the densities (black) (c) lognormal and (d) gompertz using the bandwidths h_{ucv} and h_{RT} and $\alpha = 0.5$ with sample size $n = 500$.

IV. Simulation study

A simulation study presented in Tab. I. is carried out from nonnegative life distributions. We consider the gamma, weibull, lognormal and gompertz distributions. For each target density, 100 replications of sample sizes $n = 50, 100, 500$ and 1000 are generated in order to validate the obtained results.

In terms of ISE the obtained results reveal that best results are given for moderate values of α, such as $\alpha = 0.5$ and $\alpha = 2$, compared with small and large values of α, 0.03 and 50. Besides, ISE values decrease as sample size increase. Globally UCV bandwidth selection method gives around 60% of the best results for moderate values of α. It gives the best results for gompertz distribution for all n. Except for gompertz distribution, the RT method is more appropriate than UCV for $n \geq 500$ and the opposite is true for $n \leq 100$.

Fig. 1 and Fig. 2 show plots of the gamma availability function estimator for lognormal and Gompertz target densities, for $\alpha = 0.5$ and 2 respectively. We can clearly see that the UCV method is better than RT method in the Gompertz case and that the RT method is better in the lognormal case, for $n = 500$, which confirms the simulation results in Table 1.

V. Conclusion

In this work, we have proposed a gamma kernel estimator for estimating the availability density. This estimator has good statistical properties, and simulation results have demonstrated its performance. The best results are observed for mean repair times of 0.5 and 2, and the RT method for estimating the smoothing parameter is more favorable for large sample sizes. For medium-sized samples, the UCV method can be used.

References

[1] D.W. Jacobson and S.R. Arora, "A nonexponential approach to availability modeling," *Annual Reliability and Maintainability Symposium Proceedings*, 1995, pp 253-260.

[2] D.P. Gaver and B.B. Chu, "Jackknife estimates of component and system availability," *Technometrics*, 1979, 21(4), pp 443-450.

[3] C.A. Vàsquez, V.H. Salinas-Torres and J. S. Romeo, "Bayesian estimation of the limiting availability in a repairable one-unit system," *Revista Colombiana de Estadìstica*, 2019, 42(1), pp 123-142.

[4] L.A. Baxter and L. Li, "Nonparametric estimation of the limiting availability," *Lifetime Data Analysis*, 1996, 2, pp 391-402.

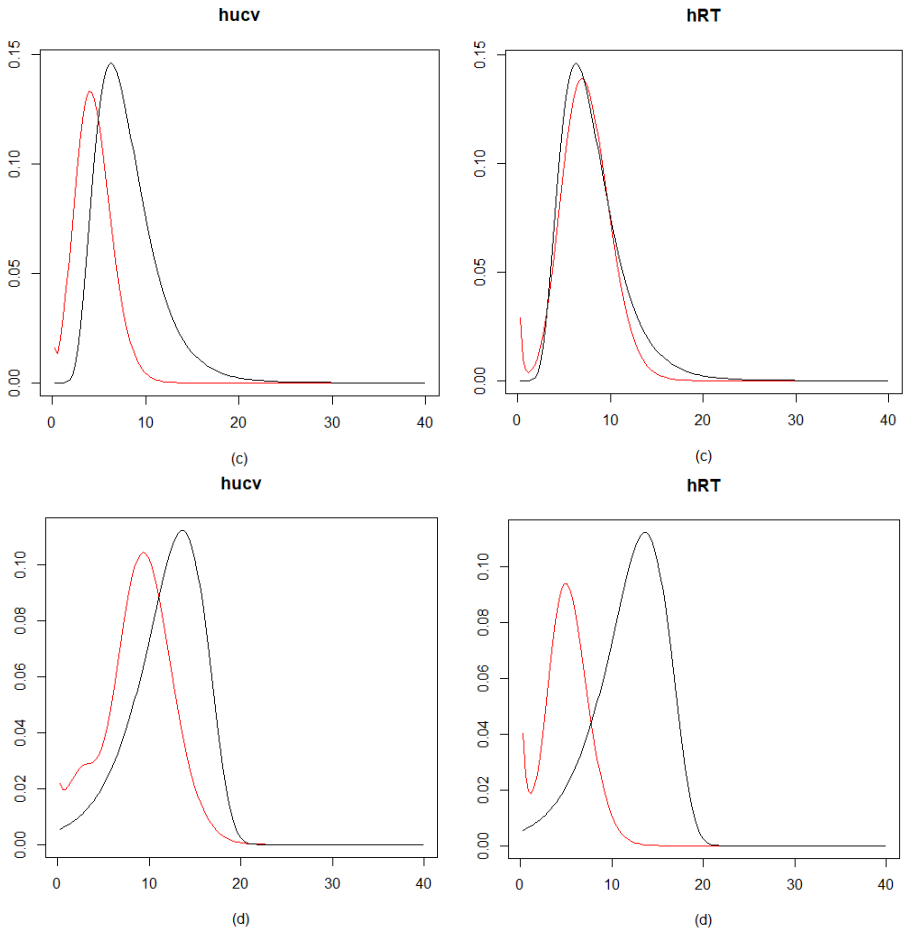


Figure 2: Plots of gamma (red) availability density function estimator for the densities (black) (c) lognormal and (d) gompertz using the bandwidths h_{ucv} and h_{RT} and $\alpha = 2$ with sample size $n = 500$.

[5] K. Huang and J. Mi, "A new non-parametric estimator for instant system availability," *Computational Statistics and Data Analysis*, 2018, 118(C), pp 18-29.

[6] S.X. Chen, "Probability density function estimation using gamma kernels," *Annals of the Institute of Statistical Mathematics*, 2000, 52(3), pp 471-480.

VI

POSTERS

Sommaire

Approches de sécurité dans les réseaux IoT

Beraza Abderrahmane
Unitée de recherche LaMOS
Abderrahmane.beraza@univ-bejaia.dz

Bouallouche Medjkoune Louiza
Unitée de recherche LaMOS
Louiza.medjkoune@univ-bejaia.dz

***Résumé*—**
L'Internet des objets (IoT) est une vision d'un Internet futur où divers objets forment un réseau. Ces objets communiquent, analysent, traitent et gèrent les données de manière autonome, utilisant les technologies RFID et les réseaux de capteurs sans fil (WSN), sans intervention humaine. Cependant, l'utilisation généralisée des objets IoT dans des domaines tels que le transport, l'éducation et la médecine pose des défis de sécurité majeurs, notamment la vulnérabilité à des attaques physiques. Des objets malveillants pourraient se faire passer pour des objets coopératifs, perturbant ainsi les communications et altérant les messages. Cet article présente les problèmes de sécurité liés à l'IoT et les travaux de recherche pour contrer de telles attaques, ainsi qu'un modèle de confiance qui permet de détecter certaines des attaques de suppression de paquets.

***Mots-Clés*—**
IoT, Internet des objets, WSN , Attaques, suppression de paquets, confiance.

I. Introduction

L'Internet des objets (IoT) est une technologie émergente qui connecte des millions d'objets physiques et virtuels via Internet, sans intervention humaine. L'objectif principal de l'IoT est de simplifier la vie en offrant divers services, tels que l'échange d'informations, la surveillance et le contrôle.

Aujourd'hui, l'IoT est largement adopté dans des domaines tels que l'agriculture, la santé, l'éducation, le transport et l'industrie, donnant lieu à des projets tels que les maisons intelligentes, les hôpitaux intelligents, les villes intelligentes, etc.

Cependant, l'IoT présente des défis de sécurité importants. Sa nature distribuée et son expansion rapide rendent certaines entités IoT vulnérables à des attaques physiques et à l'intrusion d'objets malveillants. Pour contrer ces menaces, des mécanismes de sécurité spécifiques sont nécessaires, y compris des modèles de confiance et de réputation pour surveiller les comportements des entités IoT.

Cet article abordera différentes facettes de l'Internet des objets (IoT). Nous commencerons par discuter de ses limites dans la Section 2, puis nous examinerons son architecture et sa technologie dans la Section 3. La Section 4 se penchera sur le routage des données à travers les réseaux IoT, tandis que la Section 5 traitera des problèmes de sécurité et des menaces qui y sont associés. Dans la Section 6, nous explorerons les travaux connexes et les méthodes de sécurité proposées pour contrer ces menaces. Nous effectuerons également des comparaisons entre ces techniques et les types de menaces qu'elles sont destinées à résoudre, comme détaillé dans la Section 7. Par la suite, nous nous concentrerons spécifiquement sur l'attaque de suppression de paquets et présenterons un modèle de confiance basé sur la réputation qui vise à contrer ce type d'attaque, comme détaillé dans la Section 8. Enfin, nous conclurons cet article en proposant des pistes de recherche futures.

II. Limitations des réseaux IoT

Comme mentionné précédemment, les réseaux IoT présentent certaines limitations qui rendent la tâche de sécurisation difficile. Parmi les principales limitations, on peut citer les suivantes [22] :

Contrainte énergétique : dans un environnement IoT, les objets disposent de ressources limitées en termes d'énergie et de batterie. Par conséquent, ils ne peuvent pas exécuter des algorithmes lourds et des instructions de sécurité qui pourraient rapidement épuiser leurs ressources.

Contrainte de capacité de calcul : l'IoT comprend un grand nombre de dispositifs avec des capacités de calcul et de stockage limitées, ce qui rend difficile l'application des solutions actuelles de gestion de la confiance et du calcul.

Haute scalabilité : dans les réseaux IoT, les objets sont autorisés à rejoindre ou à quitter le réseau à tout moment. Par conséquent, le système de gestion de la confiance doit résoudre ces problèmes en permettant aux nouvelles entités d'établir rapidement la confiance selon des conditions spécifiées de supervision.

Ces limitations imposent des défis significatifs en matière de sécurité dans les réseaux IoT, notamment en ce qui concerne la gestion de la confiance, la sécurité des communications et l'efficacité énergétique. Pour surmonter ces défis, des solutions spécifiques doivent être développées pour répondre aux besoins uniques de l'IoT.

III. Architecture et technologies des réseaux IoT

Les réseaux IoT se composent généralement de trois couches [1,2] : la couche Perception, la couche Réseau et la couche Application. Chacune de ces couches présente des préoccupations spécifiques en matière de sécurité. La couche Perception utilise des capteurs pour surveiller l'environnement et collecter des données, la couche Réseau gère le routage des données entre les dispositifs IoT, et la couche Application garantit l'authenticité, l'intégrité et la confidentialité des données.

Les technologies de communication clés de l'Internet des objets incluent les réseaux de capteurs sans fil (WSN) [26] et l'identification par radiofréquence (RFID) [27,28]. Les réseaux de capteurs sans fil sont utilisés pour surveiller des variables physiques telles que la température, le son et la pression, tandis que la technologie RFID agit comme un code-barres électronique pour identifier des objets ou des personnes.

IV. Routage dans les IoT

Comme pour les réseaux Ad Hoc, les réseaux IoT [9] sont constitués d'un ensemble de nœuds mobiles dynamiquement répartis, indépendants d'une infrastructure centrale. Les nœuds du réseau doivent pouvoir s'adapter aux changements de localisation dynamiques et trouver rapidement des chemins de routage dans le réseau. C'est pourquoi il est nécessaire de trouver un meilleur protocole de routage pour l'IoT. Plusieurs études ont été menées sur les protocoles existants dans les MANET (réseaux mobiles ad hoc), tels que AODV, DSR et OLSR, afin de voir la possibilité de les exploiter avec un schéma de protocole approprié pour l'IoT. Compte tenu des contraintes des objets IoT en termes de puissance de traitement, de batterie et de mémoire, un protocole de routage IPv6 [10] adapté aux réseaux à faible puissance et à pertes appelé RPL (Routing Protocol for Low-Power and Lossy Networks) a été normalisé par l'IETF en 2011 et est rapidement devenu le protocole de routage de choix pour l'IoT.

A. RPL (Routing Protocol for Low-Power and Lossy Networks)

RPL [13] est un protocole de routage IPv6 distinctif qui fonctionne avec tous les réseaux présentant des caractéristiques de faible consommation d'énergie et de pertes LLN (Low-Power and Lossy Network), tels que les réseaux de capteurs sans fil, Bluetooth et le Wi-Fi à faible puissance. RPL est un protocole proactif basé sur un algorithme de vecteur de distance qui crée une topologie de routage sous la forme d'un graphe orienté sans cycles appelé Destination Oriented Graph (DODAG) : un graphe dirigé orienté vers le nœud racine. Chaque nœud maintient ensuite plusieurs parents vers la racine, dont un seul est utilisé pour transmettre les paquets de données à la racine, tandis que les autres sont conservés en tant que routes de secours. RPL permet la communication du nœud racine avec un état de routage minimal. La topologie est créée et entretenue via deux types de paquets de contrôle appelés DIO (DODAG Information Objects) et DAO (Destination Advertisement Object) annoncés par chaque nœud. Les nœuds diffusent périodiquement des DIO sur les liens, en commençant par la racine. Les nœuds écoutent les DIO et utilisent leurs informations pour rejoindre un nouveau DODAG ou maintenir un DODAG existant. Après l'échange de DIO, chaque nœud dispose d'un ensemble de nœuds parents. Après la construction du DODAG avec les DIO, un nœud ne connaît pas ses enfants, seules les routes ascendantes sont connues. Les nœuds informent les parents de leur présence et se lient à leurs enfants avec des messages DAO.

V. Problème de sécurité dans l'IoT

En raison de la nature des réseaux IoT qui utilisent la technologie sans fil, ils sont sensibles à l'écoute indiscrète. Les objets malveillants peuvent facilement s'intégrer dans le réseau et causer un dysfonctionnement, prendre le contrôle et altérer l'intégrité des informations, causant ainsi d'importants dommages. En raison des ressources et de l'énergie limitées des nœuds IoT, les méthodes cryptographiques sont insuffisantes pour établir une confiance entre les nœuds utilisant la technologie des capteurs IoT. Les services de sécurité utilisés dans les réseaux IoT sont similaires à ceux des autres réseaux sans fil. L'objectif principal est de protéger les informations en transit en garantissant la confidentialité, l'intégrité, la disponibilité et l'authenticité des entités.

VI. Attaques et menaces dans l'Internet des Objets

Les attaques possibles visant les réseaux IoT sont similaires à celles visant d'autres réseaux utilisant des technologies de capteurs sans fil. Ces attaques exploitent une ou plusieurs failles d'un système dans le but d'atteindre un objectif spécifique, comme l'accès illégal au système, l'interruption et la perturbation d'un service, la falsification de données ou l'exploitation des ressources du système.

Modèle d'attaque :

Un attaquant peut avoir plusieurs objectifs. Il peut attaquer directement le réseau de manière malveillante pour perturber son fonctionnement, accéder au système, voler des informations ou interrompre un service. Il peut aussi être égoïste en cherchant à servir ses propres intérêts et à économiser ses ressources. D'autres attaques visent à rompre le système de confiance entre les objets du réseau. Dans cette section, nous proposons une nouvelle classification des attaques existantes dans les réseaux IoT.

A. Attaques sur les objets et l'identité

Plusieurs attaques visent directement l'objet pour prendre le contrôle de celui-ci, intercepter ses communications ou utiliser son identité. Dans un environnement IoT basé sur la confiance, chaque objet doit avoir une identité unique. Malheureusement, une entité externe peut avoir une ou plusieurs fausses identités. Parmi les attaques liées à l'objet et à l'identité, on peut citer les suivantes :

Attaque par force brute : elle consiste à essayer de casser les mots de passe et les clés de chiffrement ou d'authentification utilisés par les objets en utilisant des algorithmes qui testent un grand nombre de mots jusqu'à deviner le bon.

Attaque de cryptanalyse : l'attaquant étudie le système de cryptographie du texte chiffré afin de trouver des vulnérabilités permettant de récupérer le texte en clair à partir du texte chiffré.

Attaque d'usurpation d'identité (attaque de spoofing) : l'attaquant tente d'usurper l'identité d'un utilisateur légitime afin d'utiliser ses privilèges dans le réseau.

B. Attaques sur le réseau et les communications

Un attaquant peut utiliser des attaques classiques sur les réseaux. On peut diviser ces attaques en deux catégories : les attaques passives et les attaques actives.

1) Attaques passives :

— Surveillance des communications : l'attaquant tente d'identifier les parties de la communication qui peuvent fournir des informations, puis lance d'autres attaques.

— Écoute et analyse du trafic : il s'agit d'une attaque utilisée pour obtenir des informations sur les objets en communication et analyser la quantité de données traitées passant par le réseau.

— Attaque SYN flood : il s'agit d'un type d'attaque de déni de service (DoS). Un attaquant demande l'accès au réseau plusieurs fois, jusqu'à ce que les ressources nécessaires à chaque connexion soient épuisées ou atteignent une limite maximale.

2) Attaques actives :

— Attaque par suppression de paquets : dans cette attaque, l'attaquant supprime les paquets de données destinés à être routés. Le but de la suppression est de perturber le processus de transmission des paquets de données (comportement malveillant) ou de préserver les ressources (comportement égoïste).

— Attaque par fabrication de messages : cette attaque consiste à générer de fausses informations de routage. On peut distinguer deux types : la falsification de demandes d'erreur, qui entraîne la destruction de routes valides et donc un déni de service contre les nœuds légitimes, et la distribution de fausses routes de transmission afin de surcharger la table de routage du nœud avec des routes erronées.

— Attaques de réexécution : dans ce modèle d'attaque, un objet malveillant enregistre certains messages déjà transmis, puis les réexécute ultérieurement sans les modifier. L'objectif est d'exploiter les vulnérabilités du système.

— Attaque par déni de service (DoS) : elle se caractérise par la tentative de l'attaquant de prévenir l'utilisation légitime d'un service. Une attaque DoS peut être effectuée en inondant la cible. Les attaques DoS figurent parmi les cyberattaques les plus dangereuses et les plus efficaces contre tout type de service, car elles ne nécessitent pas d'identifier ou d'exploiter des failles de protocole ou de service, mais doivent simplement les submerger. Une attaque DoS peut paralyser le système et empêcher les objets de communiquer entre eux.

— Attaque de l'homme du milieu (Man-in-the-Middle) : l'attaquant s'interpose entre deux objets et intercepte le flux de messages de service entre eux afin de remplacer les messages par d'autres services ou de les supprimer.

C. Attaques sur le système de confiance

Dans cette catégorie d'attaques, les objets malveillants ciblent les systèmes de confiance placés entre les objets, soit en utilisant un comportement incohérent, soit en fournissant de fausses recommandations concernant les communications. Ces attaques peuvent être effectuées individuellement par un objet malveillant unique ou par collusion d'un groupe d'objets IoT. Voici les attaques sur le système de confiance dans les réseaux IoT et les nœuds capteurs :

Attaque de dénigrement (Bad mouthing) : dans ce type d'attaque, un nœud malveillant émet de fausses recommandations négatives (faux négatifs) sur un autre nœud de l'environnement IoT. Ces fausses recommandations fabriquées visent à ruiner la réputation de l'entité ciblée. Il s'agit d'une forme d'attaques de collusion en conjonction avec d'autres nœuds malveillants qui ciblent un nœud. En conséquence, le système isole le nœud victime ou réduit ses chances d'être choisi comme prestataire de services. Le nœud malveillant fournit des recommandations appropriées pour d'autres objets du réseau, apparaissant ainsi comme un recommandateur impartial aux yeux des autres entités.

Attaques de bourrage de bulletin (Ballot-stuffing) : les attaquants s'accordent pour augmenter leur réputation et celle de leurs nœuds amis en donnant de fausses recommandations positives (faux positifs). Ainsi, ils augmentent leurs chances d'être choisis comme prestataires de services.

Attaques de traîtres (Traitor attacks) : dans ces attaques, un nœud malveillant accumule une bonne réputation au départ, puis change de comportement et devient malveillant. Une fois que sa réputation chute en dessous du seuil, il modifie son comportement et se comporte de manière honnête pour augmenter sa réputation et regagner la confiance dans le réseau.

Attaques d'auto-promotion (Self-promotion attacks) : en exploitant les vulnérabilités de l'authentification et de l'intégrité des données dans l'environnement IoT, un nœud malveillant modifie sa valeur de confiance lors du processus de recommandation et fait des commentaires positifs sur lui-même. Une fois sélectionné pour un service, ce service agit de manière égoïste ou malveillante.

Attaques de blanchiment (White-washing attacks) : dans cette attaque, un nœud malveillant quitte l'environnement IoT pour laver sa mauvaise réputation. Cette attaque se

produit lorsque la valeur de confiance attribuée à la nouvelle connexion est supérieure à sa valeur de confiance actuelle. Un nœud malveillant change d'identité et revient périodiquement avec une nouvelle identité.

Attaque discriminatoire (Discriminatory attack) : dans ce type d'attaque de confiance, un nœud fournit des recommandations élevées à certains nœuds par rapport à d'autres nœuds non amicaux qui fournissent les mêmes services de qualité. Dans une autre forme, un nœud fournit un service de haute qualité à ses entités amies et non à d'autres entités.

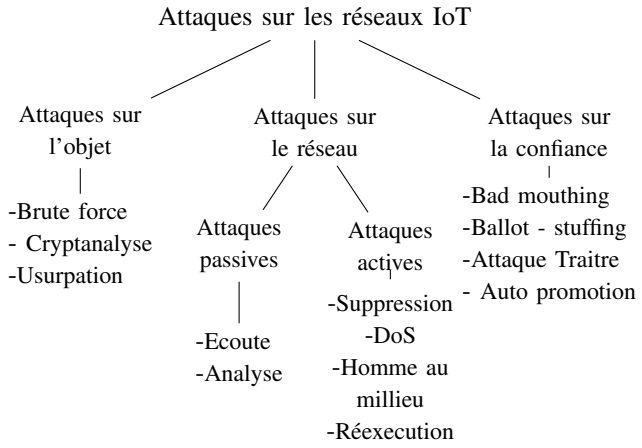


FIGURE 1. Types d'attaques dans les réseaux IoT

VII. SOLUTIONS PROPOSÉES ET TRAVAUX DANS L'INTERNET DES OBJETS (IOT)

Plusieurs approches ont été développées pour contrer les attaques sur les réseaux IoT mentionnées précédemment, en détectant les nœuds malveillants et égoïstes et en les incitant à coopérer avec d'autres nœuds du réseau. Ces approches peuvent être classées en fonction de leur nature en approches basées sur la cryptographie, approches basées sur la réputation, approches basées sur le crédit, approches basées sur la théorie des jeux et approches basées sur le clustering.

A. Approches basées sur la cryptographie

Dans le but de maintenir les objectifs de sécurité de l'IoT tels que la confidentialité, l'authentification et l'intégrité, de nombreuses mesures de sécurité cryptographique basées sur le chiffrement et le hachage des messages et des données ont été discutées. Parmi les approches proposées, citons :

Zhao et al. [16] ont présenté un schéma d'authentification mutuelle pour l'IoT, y compris les plates-formes et les objets. Le schéma repose sur le hachage et l'extraction de caractéristiques afin d'éviter les attaques par collision. Ce schéma fournit en réalité une bonne solution légère d'authentification pour l'IoT.

Porambage et coll. [17] ont proposé et conçu un protocole d'authentification avec des clés pour les réseaux de capteurs sans fil (WSN) dans une application IoT distribuée, appelé PAuthKey. Ce protocole comporte deux phases : une phase d'enregistrement pour obtenir des informations d'identification cryptographique pour les objets, et la phase d'authentification pour l'établissement de clés en communication mutuelle. En fin de compte, les utilisateurs peuvent s'authentifier directement auprès des nœuds capteurs et acquérir des données et des services. Le protocole prend en charge les applications IoT distribuées, car les certificats sont légers et peuvent être gérés par des appareils aux ressources limitées.

Salman et coll. [18] ont proposé un nouveau schéma d'authentification basé sur l'identité hétérogène qui peut être déployé à l'aide de nœuds distribués en nuage. Le processus d'authentification se compose de trois niveaux : les objets, la passerelle et le contrôleur. La première étape consiste à obtenir un certificat d'authentification pour la passerelle à partir du contrôleur. La deuxième phase est l'enregistrement des objets auprès de leur passerelle. La dernière phase est l'authentification des objets près de la passerelle. Le schéma proposé est immunisé contre les attaques d'usurpation d'identité, les attaques de l'homme du milieu et les attaques de réexécution.

B. Approches basées sur la réputation

L'une des méthodes les plus importantes pour détecter les attaques et les comportements non coopératifs d'un nœud dans un réseau est basée sur la réputation. Celles-ci surveillent les nœuds lors de leurs communications et de leurs échanges de données, puis elles incitent les nœuds égoïstes et malveillants à coopérer en utilisant divers algorithmes. La réputation d'un nœud dans le réseau reflète sa confiance envers d'autres nœuds avec lesquels il souhaite communiquer. Diverses méthodes de réputation ont été proposées pour les réseaux IoT. Un modèle de gestion de la confiance a été introduit par Bordel et al. [5], basé sur deux types de réputations : implicite et explicite.

Dans leur approche, chaque nœud fait directement confiance à son cercle de nœuds voisins. La réputation implicite d'un nœud qui n'appartient pas au cercle de confiance est calculée à partir d'observations directes (communication directe) et indirectes (communications avec des nœuds appartenant au cercle de confiance du nœud). La réputation explicite d'un nœud est obtenue à partir des recommandations partagées par les nœuds appartenant au cercle de confiance. Les deux réputations sont prises en compte dans une moyenne géométrique et forment une valeur de réputation globale. Les nœuds malveillants dont la valeur de réputation est inférieure à un seuil défini sont exclus du réseau.

Chen et Chang [6] ont développé un modèle de confiance pour l'Internet des objets basé sur la réputation des capteurs WSN. Ce modèle établit une confiance entre les nœuds locaux et globaux en utilisant deux niveaux de réputation.

La confiance locale est évaluée en surveillant directement les communications des nœuds voisins, enregistrant les transactions réussies au fil du temps. Cette réputation directe est susceptible de changer. Pour obtenir une confiance plus précise, ils ont introduit une réputation globale en combinant la réputation directe d'un nœud avec les réputations directes partagées par d'autres nœuds dans le réseau, permettant une meilleure évaluation de la confiance envers un nœud donné.

Chen et al. [7] ont conçu un système de sécurité pour l'Internet des objets basé sur la gestion de la réputation entre les utilisateurs du réseau. Ce système calcule la réputation d'un nœud en utilisant la satisfaction directe, qui résulte de l'historique des interactions passées entre les nœuds et des recommandations fournies par d'autres utilisateurs. Le modèle définit trois types de relations : Amitié, contact social et communauté d'intérêt. Chaque utilisateur stocke les listes de ses relations via un appareil high-tech, permettant aux autres objets de la même personne d'y accéder. Les relations entre utilisateurs facilitent la mesure des recommandations de confiance en se basant sur les similarités, afin de combiner la confiance directe et indirecte dans un nœud et ainsi obtenir sa réputation globale. Ce système permet de détecter les nœuds malveillants qui cherchent à perturber le système de confiance en fournissant de fausses informations, telles que recommander des nœuds malveillants ou s'auto-recommander, ou encore en endommageant la réputation d'autres nœuds par de fausses recommandations.

Dans [8], Kokoris-Kogias et al. proposent un modèle hybride de confiance pour l'Internet des objets social IoT. Ce modèle repose sur le calcul de l'indice de confiance des nœuds voisins en fonction des observations et des expériences des interactions directes. Un nœud (Objet) souhaitant demander un service à un nœud donné ayant le service demande d'abord sa valeur de réputation aux autres nœuds appartenant à son cercle de confiance. Si les informations sur la réputation du nœud de destination ne sont pas fournies ou ne sont pas suffisantes, le nœud demandeur du service consulte une plateforme centrale qui contient une vue plus large du réseau et qui stocke l'historique des valeurs de réputation et des interactions entre chaque nœud. Le nœud décide ensuite en fonction des degrés de réputation de faire confiance au nœud de destination ou de le rejeter.

Xiao et al. [10] ont proposé un modèle de confiance basé sur la garantie et la réputation pour les réseaux sociaux IoT. Les objets sont connectés et font confiance directement à leur passerelle du réseau. Le modèle utilise deux paramètres qui sont le crédit et la réputation. Chaque nœud souhaitant obtenir un service paie une commission en crédits au bénéfice du nœud fournissant le service si ce dernier fournit correctement le service. Sinon, si le nœud refuse de coopérer ou ne transmet pas correctement le service, il devra payer des crédits en commission. Un serveur de réputation stocke et met à jour la valeur de réputation de chaque objet en fonction de ses transactions. Ce serveur partage les valeurs de réputation avec les nœuds souhaitant communiquer dans le réseau.

C. Approches basées sur la théorie des jeux

La théorie des jeux propose un modèle mathématique pour la compétition entre les nœuds et ainsi les inciter à coopérer et à obtenir la meilleure performance du jeu. Il s'agit de l'une des méthodes économiques utilisées en informatique et qui peut fournir une analyse globale entre les communications des objets dans un réseau IoT.

Nobahary et al. [19] ont proposé une méthode hybride qui tire parti des méthodes basées sur la réputation et de celles basées sur la théorie des jeux pour détecter et stimuler les nœuds non coopératifs. La stratégie attribue une valeur de réputation gagnée à tous les nœuds tout en jouant au jeu. Les nœuds ayant une mauvaise réputation ne peuvent pas être actifs et essaient donc d'envoyer autant de paquets de données que possible pour gagner plus de réputation. Les nœuds souhaitant gagner en réputation doivent donc coopérer avec d'autres nœuds. Tant que la réputation d'un nœud est supérieure à un certain seuil, le nœud a la possibilité de coopérer avec d'autres nœuds. Ce schéma réduit la consommation d'énergie et augmente le débit du réseau en réduisant le retour des paquets de données. L'approche permet de détecter les nœuds égoïstes et malveillants dans le réseau tout en réduisant le taux de faux positifs et de faux négatifs.

Les auteurs de [23] nous présentent une nouvelle stratégie légère de confiance (EETE) qui utilise la théorie des jeux pour améliorer la sécurité de l'IoT facilitée par des capteurs regroupés. La méthode prend en compte deux facteurs : la fiabilité et l'efficacité énergétique, tous deux essentiels pour la présence d'un capteur IoT actif dans un environnement non protégé. EETE utilise la théorie des jeux évolutifs pour former des clusters et la théorie des jeux non coopératifs pour détecter les nœuds réseau non fiables afin de permettre une communication sûre et efficace sur le plan énergétique entre les nœuds. Pour détecter les nœuds malveillants, les modèles de jeu limitent également les diffusions superflues. Grâce à l'efficacité énergétique et à la notation de confiance, ainsi qu'au temps d'évaluation, la méthode EETE améliore l'identification des nœuds malveillants. Cependant, les attaques externes telles que les attaques par déni de service (DoS), les attaques de trou noir et les attaques de trou de ver ne sont pas détectées.

Dans [24], Wu et Wang ont développé une approche de sécurité basée sur la théorie des jeux. Les deux joueurs dans ce type de jeu sont d'abord définis comme un attaquant et une défense. En tant qu'ensembles stratégiques de l'attaquant et du défenseur, les ressources d'attaque allouées aux nœuds du réseau et un seuil de détection commun sont choisis. Ensuite, l'existence, l'unicité et le calcul du modèle de jeu dans la situation du consensus complet avec un nombre illimité d'itérations sont étudiés. Il est également expliqué comment

choisir le meilleur nombre d'itérations. La relation entre les équilibres de Nash des modèles de jeu de consensus complet et incomplet est étudiée quantitativement. Cette approche peut identifier les attaques par déni de service distribué (DDoS) et éviter les problèmes liés aux services de réseau de l'Internet des objets.

D. Approches basées sur le clustering

Nobahary et al. [20] proposent un nouveau schéma appelé DISOT qui permet de détecter les nœuds égoïstes. Le mécanisme proposé utilise la méthode de clustering pour fournir une surveillance où tous les nœuds sont regroupés. Le réseau est composé de plusieurs clusters, chacun ayant un chef de cluster responsable de la surveillance de ses nœuds membres. Le chef de cluster a une communication directe avec les chefs des autres clusters multi-sauts, un chef de cluster principal surveille l'activité des autres chefs de cluster. L'algorithme proposé se compose de trois phases : la phase de configuration et de regroupement, la phase globale et la phase locale. Au cours de la première phase, les clusters sont formés entre les nœuds voisins, puis un chef de cluster sera déterminé pour chaque groupe. Dans la deuxième phase (phase globale), le chef de cluster principal, qui communique avec les autres chefs de cluster, surveille et signale la présence de comportements égoïstes sans identifier le nœud concerné. La phase locale débutera alors au niveau des clusters où le comportement égoïste a été détecté afin d'identifier précisément le nœuds égoïstes responsable de ce comportement. Les nœuds égoïstes peuvent ensuite être isolés du réseau.

Les auteurs de [25] ont proposé IoT-TM, une nouvelle approche de gestion de la confiance pour l'Internet des objets basée sur le clustering. La méthode repose sur des algorithmes pour assurer que le réseau continue à évoluer. Le premier élimine les valeurs qui ne sont pas cohérentes avec les valeurs de confiance pour éviter les attaques de diffamation et s'assurer que seules les bonnes valeurs de confiance pour un service IoT sont prises en compte. Le deuxième méthode, qui repose sur la valeur de confiance entre les nœuds, crée des clusters intelligents. Les nœuds IoT migrent d'un cluster à un autre en fonction de la confiance dans la troisième méthode. Enfin, un dernier algorithme examine les états actuels des nœuds du cluster IoT par rapport aux valeurs de confiance pour choisir quel cluster rejoindre.

Nous avons d'abord étudié les différentes attaques qui existent dans les réseaux IoT, puis nous avons classé les approches de sécurité que nous avons trouvées dans la littérature visant à améliorer la sécurité et la confiance entre les nœuds d'un réseau IoT. Dans cette section, nous allons comparer les approches mentionnées ci-dessus en fonction des types d'attaques auxquelles elles font face. Le tableau Tab.1 ci-dessous présente une comparaison entre les approches citées selon les types d'attaques auxquelles elles répondent : En analysant Tab. 1, nous constatons que les approches basées sur la cryptographie visent à renforcer les réseaux IoT contre les menaces liées aux objets et aux identités, mais elles ne sont pas efficaces pour contrecarrer les objets malveillants ou égoïstes qui s'introduisent dans le réseau. Les approches basées sur la réputation, le crédit, la théorie des jeux et les approches en cluster permettent de gérer la confiance et permettent la détection et l'isolation des objets non coopératifs qui ne se comportent pas correctement et nuisent au bon fonctionnement du réseau.

TABLE I. Approches et types d'attaques.

Type of approache	Approaches	Brute force	fabrication messages	Spoofing	Replay	Man in the middle	Drop packets	Trust/ attacks	DoS
Cryptography	Zhao & al [16]	X		X	X	X			
	Porambage & coll [17]	X		X	X	X			
	Salman & coll [18]								
Reputation	Bordel & al [5]		X		X	X		X	
	Chen & Chang [6]						X	X	
	Chen & al [7]							X	
	Kokoris-Kogias & al [8]							X	
	Xiao & al [10]							X	
Game theory	Nobahary & al[19]						X	X	
	Rani & al [23]							X	
	Wu & al [24]							X	X
Cluster	Nobahary & al [20]						X	X	
	Alshehri & al [25]							X	

VIII. ATTAQUES DE SUPPRESSION DES PAQUETS

Nous nous sommes intéressés par les attaques existantes sur le réseau, et plus précisément sur les protocoles de routage et de communication, où un attaquant refuse de coopérer en supprimant les paquets de données à acheminer. On peut trouver deux types d'attaques de suppression de paquets qui sont :

A. Attaque du trou noir (Black Hole)

Les attaques du trou noir sont des attaques visant la topologie du réseau, où un nœud malveillant appartenant au réseau laisse tomber les paquets qu'il est censé transmettre. Dans les reseaux IoT, une attaque du trou noir, comme illustré dans la Fig. 2 ci-dessous, peut être passive ou active. Un attaquant passif ayant un comportement égoïste laisse tomber tous les paquets de contrôle qu'il reçoit. Cela peut forcer un nœud à se connecter au réseau via un chemin moins optimal, voire l'empêcher de rejoindre le réseau s'il ne peut pas trouver un autre chemin à travers d'autres nœuds. Ces attaques passives du trou noir n'affectent pas le fonctionnement du réseau et ne peuvent pas provoquer d'isolement topologique.

Un attaquant actif ayant des intentions malveillantes ne rejette aucun message de contrôle reçu et assure la connectivité au nœud destinataire. Cependant, il rejette tous les paquets de données reçus. Il peut également laisser tomber ses propres paquets pour éviter d'être détecté. Cette attaque réduit le taux de livraison des paquets du réseau et peut retarder indéfiniment les paquets provenant des nœuds affectés.

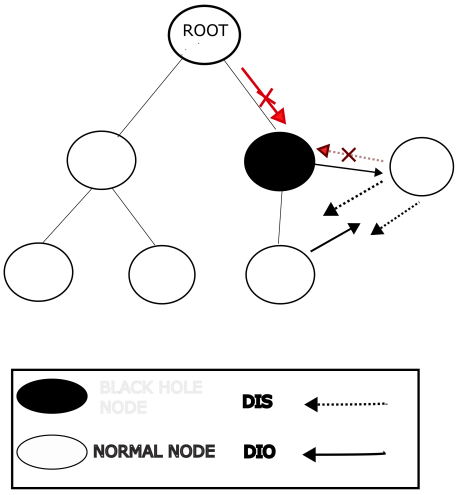


FIGURE 2. Black hole attack in RPL

B. Attaque de suppression sélective

Contrairement à l'attaque du trou noir où le nœud malveillant laisse tomber tous les paquets, l'attaque de suppression sélective [35] consiste à sélectionner les paquets à transmettre et à laisser tomber quelques paquets provenant de nœuds internes spécifiques. Cette attaque est difficile à détecter car il peut y avoir plusieurs raisons pour lesquelles le paquet est perdu. Un nœud malveillant peut sélectionner les paquets à transmettre tout en maintenant sa fiabilité, éliminant ainsi les soupçons quant à des erreurs dans les protocoles de routage. L'objectif de l'attaque est de perturber les chemins de routage et de filtrer tout protocole.

C. Travaux existants

Dans [29], les auteurs ont proposé un système de réputation en chaîne avec un seuil adaptatif nommé CRS-A pour identifier les réseaux de capteurs sans fil (WSN) vulnérables aux attaques de transfert sélectif. CRS-A évalue le comportement progressif en calculant la différence entre la perte de paquets typique estimée et la perte de paquets observée. Pour améliorer la précision de la suppression des paquets, les auteurs ont développé une méthode probabiliste pour déterminer le meilleur seuil d'évaluation. CRS-A démontre une précision élevée dans la suppression de paquets avec de faibles probabilités de détection erronée.

Une méthode basée sur la confiance pour le protocole RPL a été proposée dans [30]. Elle repose sur le calcul d'une valeur de confiance pour chaque nœud du réseau RPL et utilise ces valeurs pour prendre des décisions de routage. Ainsi, cette approche combine les avantages de choix de routage optimaux tout en isolant efficacement les attaques de type trou noir et le transfert sélectif. La réputation de chaque nœud est calculée en fonction du ratio de paquets envoyés par un autre nœud par rapport aux paquets reçus par ce même nœud à travers lui. Les résultats de simulations montrent que cette approche offre des avantages significatifs en termes de performances du réseau et de sécurité.

Dans [31], les auteurs proposent une approche pour détecter les nœuds trou noir qui suppriment les paquets de données dans l'Internet des objets en utilisant le protocole de routage RPL. Pour ce faire, l'approche est basée sur l'estimation du temps d'arrivée des paquets au nœud récepteur par d'autres nœuds LLN tout en surveillant les changements dans la topologie DODAG. Les sous-arbres qui exposent un nœud à une attaque de type trou noir sont alors identifiés. Un appel de détection est ensuite émis pour explorer les routes potentiellement menaçantes et détecter les nœuds malveillants qui suppriment les paquets.

T CAFE [32] est une approche de sécurité basée sur la confiance pour l'Internet des objets qui calcule la confiance directe d'un nœud envers la destination. Les nœuds calculent des paramètres en temps réel. Pour calculer la valeur de confiance indirecte d'un autre nœud, un nœud utilise les recommandations de ses voisins. Le niveau de confiance final est obtenu grâce à une combinaison des méthodes directes et indirectes. L'approche basée sur la réputation proposée est un protocole qui garantit un taux de livraison élevé et une perte minimale de paquets.

Chen & Chang dans [6] proposent un modèle de confiance dans l'Internet des objets avec des capteurs WSN basé sur la réputation, appelé TRM-IoT. Ce modèle repose sur une relation de confiance entre les nœuds locaux ou globaux. La confiance locale est calculée par des observations et une surveillance directe des communications des nœuds voisins. Chaque nœud contient une table de transactions pour chacun de ses nœuds voisins, incluant des informations sur le moment et le nombre de paquets transmis avec succès. Étant donné que la réputation directe d'un nœud peut varier avec le temps, les auteurs ont incorporé une relation globale pour obtenir une valeur de confiance plus précise d'un nœud.

Pour détecter et stimuler les nœuds non coopératifs, Nobahary et al. [19] ont proposé une méthode hybride qui combine des méthodes basées sur la réputation et la théorie des jeux. Pendant le jeu, la stratégie attribue une valeur de réputation à tous les nœuds. Les nœuds avec une mauvaise réputation ne peuvent pas être actifs et tentent alors d'envoyer autant de paquets de données que possible pour améliorer leur réputation. Par conséquent, les nœuds souhaitant gagner une meilleure réputation doivent coopérer avec d'autres nœuds. Tant qu'une réputation de nœud est au-dessus d'un certain seuil, il est possible qu'il collabore avec d'autres nœuds. En réduisant le retour des paquets de données, ce schéma réduit la consommation d'énergie et augmente le débit du réseau.

Dans [33], un algorithme efficace de détection et de prévention de l'attaque trou noir a été proposé pour le protocole RPL. De plus, la découverte du trou noir a été réalisée en fonction de la valeur seuil de chaque nœud dans le

DODAG RPL. La valeur seuil a été calculée en fonction du taux de perte de paquets de chaque nœud. Les résultats ont prouvé l'efficacité du système proposé en termes de taux de découverte des attaques, de détection de bout en bout et de taux de livraison des paquets.

Les auteurs dans [34] ont proposé une stratégie de gestion de la confiance pour sécuriser la topologie du routage réseau dans le cadre de RPL. Les nœuds analysent s'ils doivent mettre à jour leurs parents DAG préférés ou rejoindre la structure de routage DAG pour envoyer des paquets de données. Le Trustee (ou fiduciaire), en revanche, est l'entité examinée qui se substitue à l'entité potentielle qui sera sélectionnée en tant que parent préféré. Cette connexion résulte d'interactions directes et d'observations, ou de "confiance directe", ainsi que de recommandations données et reçues des nœuds voisins, ou de "confiance indirecte". Les résultats se sont révélés efficaces et pertinents pour identifier et bloquer les nœuds hostiles qui lancent des attaques de type trou noir.

Discussion

Après avoir étudié quelques approches qui visent à détecter et à isoler l'attaque de suppression de paquets dans les réseaux IoT, on trouve que la plupart d'entre elles utilise le principe de réputation dans le calcul du degrés de confiance entre des noeuds. La plupart de ses approches montre une efficacité et améliore le taux de livraison de paquets. Néanmoins, presque toutes les approches de réputation existantes, ne peuvent pas détecter les nœuds malveillants lorsqu'ils lancent une attaque de suppression sélective, ce qui est l'une des limitations importantes. Ces nœuds suppriment des paquets de données à un faible taux dans le but de maintenir leur valeur de réputation au-dessus du seuil toléré et d'éviter d'être considérés comme des nœuds malveillants, ce qui crée une injustice envers les nœuds coopératifs qui coopèrent pleinement.

IX. Approche proposée

Dans cette section, nous présentons une modèle d'approche afin de contrer l'attaque de suppression sélective. Ce dernier sera organisé autour de trois modules : la surveillance, la réputation et l'exclusion.

A. Module de surveillance

Le module de surveillance dans le contexte du protocole RPL est responsable de suivre les activités de transfert de paquets de données par les nœuds voisins. Contrairement à l'approche du "chien de garde", RPL est conçu pour être économe en énergie et adapté aux réseaux à faible consommation d'énergie et à faible bande passante. Par conséquent, nous devons adapter l'idée de surveillance aux principes de RPL.

Dans le protocole RPL, chaque nœud maintient un voisinage de nœuds parents et enfants pour acheminer les paquets de données. Le nœud parent est responsable du transfert des paquets vers le nœud racine du réseau. Le nœud enfant est le nœud qui transfère les paquets vers un nœud parent. Le module de surveillance dans RPL pourrait être conçu pour surveiller la fiabilité des nœuds parents. Par exemple, chaque nœud enfant pourrait suivre le taux de succès des transmissions de ses nœuds parents en surveillant les messages de contrôle et les paquets de données. Si un noeud parent transmet correctement les paquets de données, le noeud enfant enregistre un évènement positif sur ce dernier. Sinon si le noeud parent ne parvient pas à acheminer correctement les paquets de données, le nœud enfant enregistre un évènement négatif.

B. Module de réputation

Le module de réputation est responsable de la gestion des valeurs de réputation des nœuds dans le processus de transfert de données en fonction des événements enregistrés par le module de surveillance. La valeur de réputation est augmentée si le module de surveillance détecte un événement positif, tandis qu'elle est diminuée en cas de détection d'un événement négatif. La plupart des approches de réputation existantes utilisent les mêmes taux d'augmentation et de diminution pour mettre à jour les valeurs de réputation des nœuds. En suivant cette approche, les nœuds avec une réputation élevée et ceux avec des valeurs de réputation faibles sont traités de manière égale lorsqu'ils sont impliqués dans un événement, ce qui conduit à une injustice envers les nœuds à réputation élevée qui coopèrent pleinement.

Cependant, nous proposons de varier le taux d'augmentation et de diminution de la réputation d'un nœud en fonction de ses valeurs de réputation passées. Ainsi, les nœuds seront traités différemment lorsqu'ils seront impliqués dans des événements positifs et négatifs.

C. Module d'exclusion

Le module d'exclusion assume la responsabilité de sanctionner les nœuds malveillants et de les isoler du réseau. Lorsqu'un nœud détecte qu'un de ses nœuds parents a une réputation inférieure au seuil défini, il considère ce nœud parent comme malveillant et l'ajoute à la liste des nœuds malveillants détectés. Ensuite, le nœud partage avec les autres nœuds un rapport de malveillance pour l'informer de la détection du nœud malveillant. Ce rapport sera ensuite routé vers la racine du réseau, qui sera chargée à son tour de diffuser l'information et d'isoler les communications passant par le nœud malveillant.

X. Conclusion

Notre étude a examiné en détail les problèmes de sécurité liés à l'Internet des objets (IoT) et les différentes approches visant à contrer ces menaces tout en établissant la confiance entre les objets connectés. Nous avons identifié les types potentiels d'attaques ciblant les réseaux IoT et classé les approches existantes en fonction de leur capacité à contrer ces attaques.

Ensuite, nous avons examiné de manière plus précise l'attaque de suppression de paquets ainsi que les approches existantes qui visent à contrecarrer ce type d'attaques. Nous avons

établi un modèle de réputation qui pourra venir à améliorer la détection de manière plus efficace des nœuds malveillants au sein des réseaux IoT. Cette recherche contribuera à renforcer la sécurité de ces réseaux en constante évolution et à garantir leur utilisation sécurisée dans divers domaines d'application.

Pour la suite de notre travail, nous envisageons d'implémenter notre modèle de réputation sur un simulateur des réseaux IoT, puis comparer nos résultats avec d'autres approches de références.

RÉFÉRENCES

[1] Madakam, S., Lake, V., Lake, V., Lake, V. (2015). Internet of Things (IoT) : A literature review. Journal of Computer and Communications, 3(05), 164.

[2] Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M. (2013). Internet of Things (IoT) : A vision, architectural elements, and future directions. Future generation computer systems, 29(7), 1645-1660.

[3] Mahmoud, R., Yousuf, T., Aloul, F., Zualkernan, I. (2015, December). Internet of things (IoT) security : Current status, challenges and prospective measures. In 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST) (pp. 336-341). IEEE.

[4] Buttyan, L., & Hubaux, J. P. (2001). Nuglets : a virtual currency to stimulate cooperation in self-organized mobile ad hoc networks(No. LCA-REPORT-2001-011).

[5] Bordel, B., Alcarria, R., De Andrés, D. M., You, I. (2018). Securing Internet-of-Things systems through implicit and explicit reputation models. IEEE Access, 6, 47472-47488.

[6] Chen, D., Chang, G., Sun, D., Li, J., Jia, J., Wang, X. (2011). TRM-IoT : A trust management model based on fuzzy reputation for internet of things. Computer Science and Information Systems, 8(4), 1207-1228.

[7] Chen, R., Guo, J., Bao, F. (2014). Trust management for SOA-based IoT and its application to service composition. IEEE Transactions on Services Computing, 9(3), 482-495.

[8] Kokoris-Kogias, E., Voutyras, O., Varvarigou, T. (2016, September). TRM-SIoT : A scalable hybrid trust reputation model for the social Internet of Things. In 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA) (pp. 1-9). Ieee.

[9] Zhong, S., Chen, J., & Yang, Y. R. (2003, March). Sprite : A simple, cheat-proof, creditbased system for mobile ad-hoc networks. In INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies (Vol. 3, pp. 1987-1997). IEEE.

[10] Xiao, H., Sidhu, N., Christianson, B. (2015, August). Guarantor and reputation based trust model for social internet of things. In 2015 international wireless communications and mobile computing conference (IWCMC) (pp. 600-605). IEEE.

[11] Ahmed, A. I. A., Ab Hamid, S. H., Gani, A., Khan, M. K. (2019). Trust and reputation for Internet of Things : Fundamentals, taxonomy, and open Research Challenges. Journal of Network and Computer Applications, 145, 102409.

[12] Hammi, M. T., Hammi, B., Bellot, P., Serhrouchni, A. (2018). Bubbles of Trust : A decentralized blockchain-based authentication system for IoT. Computers Security, 78, 126-142.

[13] Mohamed Tahar Hammi. Sécurisation de l'Internet des objets. Réseaux et télécommunications [cs.NI]. Université Paris-Saclay, 2018. Français. NNT : 2018SACLT006.tel-01997261

[14] Chen, R., Bao, F., Guo, J. (2015). Trust-based service management for social internet of things systems. IEEE transactions on dependable and secure computing, 13(6), 684-696.

[15] Deogirikar, J., Vidhate, A. (2017, February). Security attacks in IoT : A survey. In 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 32-37). IEEE.

[16] G. Zhao, X. Si, J. Wang, X. Long, and T. Hu, "A novel mutual authentication scheme for Internet of Things," in Int'l Conference on Modelling, Identification and Control (ICMIC), 563-566, 2011.

[17] P. Porambage, C. Schmitt, P. Kumar, A. Gurtov et M. Ylianttila, ≪Pauthkey : Un protocole d'authentification omniprésent et un schéma d'établissement de clé pour les réseaux de capteurs sans fil dans les applications IOT distribuées≫, dans Journal international des réseaux de capteurs distribués, 2014.

[18] O. Salman, S. Abdallah, IH Elhajj, A. Chehab et A. Kayssi, ≪Identity-based authentication scheme for the Internet of Things≫, dans Symposium IEEE 2016 sur les ordinateurs et la communication (ISCC), Juin 2016, p. 1109-1111.

[19] Nobahary, S., Garakani, H. G., Khademzadeh, A., & Rahmani, A. M. (2019). Selfish node detection based on hierarchical game theory in IoT. EURASIP Journal on Wireless Communications and Networking, 2019(1), 1-19.

[20] Nobahary, S., Gharaee Garakani, H., Khademzadeh, A., & Rahmani, A. M. (2018). DISOT : Distributed Selfish Node Detection in Internet of Things. International Journal of Information and Communication Technology Research, 10(3), 19-30.

[21] Bansal, S., & Baker, M. (2003). Observation-based cooperation enforcement in ad hoc networks. arXiv preprint cs/0307012.

[22] Yang, Y., Wu, L., Yin, G., Li, L., & Zhao, H. (2017). A survey on security and privacy issues in Internet-of-Things. IEEE Internet of Things Journal, 4(5), 1250-1258.

[23] Rani, R., Kumar, S., & Dohare, U. (2019). Trust evaluation for light weight security in sensor enabled Internet of Things : Game theory oriented approach. IEEE Internet of Things Journal, 6(5), 8421-8432.

[24] Wu, H., & Wang, W. (2018). A game theory based collaborative security detection method for Internet of Things systems. IEEE Transactions on Information Forensics and Security, 13(6), 1432-1445.

[25] Alshehri, M. D., Hussain, F. K., & Hussain, O. K. (2018). Clustering-driven intelligent trust management methodology for the internet of things (CITM-IoT). Mobile networks and applications, 23(3), 419-431.

[26] Hussain, S., Schaffner, S., & Moseychuck, D. (2009, May). Applications of wireless sensor networks and RFID in a smart home environment. In 2009 Seventh Annual Communication Networks and Services Research Conference (pp. 153-157). IEEE.

[27] Juels, A. (2006). RFID security and privacy : A research survey. IEEE journal on selected areas in communications, 24(2), 381-394.

[28] Sun, C. (2012). Application of RFID technology for logistics on internet of things. AASRI Procedia, 1, 106-111.

[29] Ren, J., Zhang, Y., Zhang, K., Shen, X. (2016). Adaptive and channel-aware detection of selective forwarding attacks in wireless sensor networks. IEEE Transactions on Wireless Communications, 15(5), 3718-3731.

[30] Airehrour, D., Gutierrez, J., Ray, S. K. (2017). A trust-aware RPL routing protocol to detect blackhole and selective forwarding attacks. Journal of Telecommunications and the Digital Economy, 5(1), 50-69.

[31] Sahay, R., Geethakumari, G., Mitra, B., Thejas, V. (2018, December). Exponential smoothing based approach for detection of blackhole attacks in IoT. In 2018 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS) (pp. 1-6).

[32] Kandhoul, N., Dhurandher, S. K., Woungang, I. (2019). T CAFE : a trust based security approach for opportunistic IoT. IET Communications, 13(20), 3463-3471.

[33] Neerugatti, V., Reddy, A. R. M., Rama, A. (2018). Detection and prevention of black hole attack in RPL protocol based on the threshold value of nodes in the internet of things networks. Int. J. Innov. Technol. Explor. Eng, 8(9).

[34] Lahbib, A., Toumi, K., Elleuch, S., Laouiti, A., Martin, S. (2017, October). Link reliable and trust aware RPL routing protocol for Internet of Things. In 2017 IEEE 16th International Symposium on Network Computing and Applications (NCA) (pp. 1-5). IEEE.

[35] Pongle, P., Chavan, G. (2015, January). A survey : Attacks on RPL and 6LoW-PAN in IoT. In 2015 International conference on pervasive computing (ICPC) (pp.1-6).

Semi-parametric estimation of the hazard rate function using the Champernowne transformation

Bourouina Massilva
Research Unit LaMOS,
University of Bejaia, Algeria
massilva.bourouina@univ-bejaia.dz

Bareche Aicha
Research Unit LaMOS,
Faculty of Technology,
University of Bejaia, Algeria
aicha.bareche@univ-bejaia.dz

Ziane Yasmina
Research Unit LaMOS,
University of Bejaia, Algeria
yasmina.ziane@univ-bejaia.dz

***Abstract*—**
The objective of this work is to use another important alternative to parametric and nonparametric methods called semi-parametric approach, to estimate the hazard rate function. To do this, we effectuate the state of the art in semi-parametric estimation of the density function f and the hazard rate function λ. Then we propose a new estimator for the hazard rate function in the semi-parametric case using the Champernowne transformation.

***Keywords*—**
density, hazard rate, semi-parametric estimation, Champernowne transformation, Beta kernel.

I. Introduction

Among the approaches used in finance and insurance to model losses are parametric and nonparametric approaches. The initial one consists of assuming that the density f belongs to a family of distributions with a finite number of parameters, but this estimating method has its drawbacks: choosing a good parametric model, identifying the best way to estimate the parameters and identifying the threshold between large and small losses. To remedy this problem, the second approach has emerged, since it does not require the specification of a parametric model, this estimation method involves estimating the entire unknown density function.

There are various nonparametric methods for estimating the density function, one can cite: the histogram estimation method (see Abou-Jouadé [1] and Geffroy [2]), the orthogonal series estimation method (see Cencov [3], Schwartz [4], Kronmal and Tarter [5], Wahba [6], Bosq [7], Saadi and Adjabi [8]), the spline interpolation method (see [9]) and the kernel method which is the most used for its flexibility and simplicity (see [10] and [11]). Although this method solves the problem of parametric model specification, it has one major drawback, which is the problem of edge bias that arises from the allocation of weights by the fixed symmetric kernel outside the distribution support when smoothing near the boundary is performed which gives a non-consistent estimator. To avoid this problem, asymmetric kernel estimators are used (see for example Hirukawa and Sakudo [12], Makhloufi et al. [13] and Marchant et al. [14]).

Kokonendji and Libengué [15] have proposed a method for constructing continuous associated kernels, which principle is called the dispersion-mode that may be summarized as follows : set the target x to the mode and the smoothing parameter h to the dispersion parameter.

Another interesting alternative to parametric and nonparametric methods is the semi-parametric approach, which is a combination of both approaches, and is suitable for heavy-tailed distributions. It has the advantage of producing a density function that can be adapted to the whole range of data, without leaving out any information about the sample (small, medium and large). in other words, we use the parametric approach, employing a parametric estimator to transform the initial data set, then we apply the classical kernel density estimator proposed by Parzen [11] and Rosenblatt [10] to the resulting sample.

Semi-parametric estimation emerged to demonstrate that the kernel density estimator could be improved by transforming the data set with a few specific transformations and reducing data skewness (see for example : Bolancé et al. [16], Bolancé [17], Buch-Larsen et al. [18], Gustafsson et al. [19], Balosooriya and Low [20]).

Chekkal et al. [21] have estimated the hazard rate function (HR) in the context of positively skewed data using the nonparametric approach and the Generalized Birnbaum-Saunders (GBS) kernel family.

The HR function noted λ is defined by the following formula:

$$\lambda(x) = \frac{f(x)}{1 - F(x)}, \quad x > 0, \tag{1}$$

where f is unknown pdf of a random variable X, and F its cumulative distribution function (cdf).

There is currently only one work on semi-parametric estimation of the hazard rate function, carried out by Hua et al. [22]. The idea of the latter, is to develop an

estimator composed of two parts, the first is parametric and the second represents the nonparametric correction function in the case of density. This estimator is then employed as a foundation for estimating the hazard rate. Hence, inspired by the work of Buch-Larsen et al. [18], the idea behind our work consists in applying the semi-parametric approach to propose a new estimator for the hazard rate function, but this time by applying the Champernowne distribution as a parametric start and the Beta kernel for the nonparametric estimation. A suitable algorithm is elaborated for further numerical issues.

II. The semi-parametric approach in the case of the density function

The semi-parametric approach consists firstly of using a parametric distribution in our case we use the generalized Champernowne distribution and then applying the asymmetric Beta kernel to the resulting sample. This choice of kernel is made because the new data obtained by the Champernowne transformation are defined on the $[0,1]$ support.

A. Champernowne generalized distribution

The Champernowne generalized distribution was used in Buch-Larsen et al. [18] when modeling insurance claims. Its cdf is defined for $x \geq 0$ and has the form:

$$T_{\alpha,M,c}(x) = \frac{(x+c)^{\alpha} - c^{\alpha}}{(x+c)^{\alpha} + (M+c)^{\alpha} - 2c^{\alpha}}, \forall x \in R_{+}, \quad (2)$$

where $\alpha > 0, M > 0$ and $c \geq 0$.

Its density function is defined by the following formula:

$$t_{\alpha,M,c}(x) = \frac{\alpha(x+c)^{\alpha-1}((M+c)^{\alpha} - c^{\alpha})}{((x+c)^{\alpha} + (M+c)^{\alpha} - 2c^{\alpha})^2}, \forall x \in R_{+}. \quad (3)$$

• For the Champernowne distribution, we have: $T_{\alpha,M,c}(M) = 0.5$. In this case M is estimated as the median of the data set.

We also have to estimate the pair (α, c) which maximizes the log likelihood function defined as follows:

$$l(\alpha, c) = N \log \alpha + N \log((M+c)^{\alpha} - c^{\alpha}) +$$

$(\alpha - 1)\sum_{i=1}^{N} \log(X_i + c) - 2\sum_{i=1}^{N} \log((X_i + c)^{\alpha} + (M + c)^{\alpha} - 2c^{\alpha})$.(4) After estimating these parameters, the next step is to apply a nonparametric estimator to the resulting sample. The form of the kernel estimator will be defined in the next subsection.

B. Kernel method

Let $X_1, \ldots, X_n$ un $n-$sample from a random variable X, with unknown density function f. The classical kernel estimator proposed by Rosenblatt [10] and Parzen [11] is defined by the following formula:

$$\widehat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K(\frac{x - X_i}{h_n}), \quad (5)$$

where: K is a kernel and h_n is the smoothing parameter.

1) Choice of kernel and smoothing parameter: There are various methods for estimating the smoothing parameter h. The first approach consists in minimizing the integrated mean square error MISE, but the disadvantage of this method is that the smoothing parameter obtained always depends on unknown quantities (see Scott [23] and Silverman [24]). The second class represents cross-validation methods, one can cite:
unbiased cross-validation UCV (see Rudemo [25] and Bowman [26]), biased cross-validation BCV (see Scott and Terrel [27]), and the smoothed cross-validation approach (see Hall et al. [28]). The third approach is also called the plug-in method, which involves replacing the unknown quantities of the smoothing parameter with other selected quantities (see Park and Marron [29], Sheather and Jones [30]). Another method of interest is the Bayesian approach (see Zougab et al. [31]).
For the choice of the kernel K, and when the density function is defined on the bounded support or bounded on one side, the disadvantage of using a symmetric kernel is that it assigns a weight outside the support when smoothing near the edge is performed, which causes the problem of boundary bias and gives a non-consistent estimator. To remedy this problem, we need to use continuous asymmetric kernels. Kokonendji and Libengué proposed a technique for constructing continuous associated kernels called: the mode-dispertion principle (see [15]). Estimators with a continuous asymmetric associated kernel are of the following form:

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x,h}(X_i), \quad (6)$$

where: $K_{x,h}$ is the asymmetric kernel, x is the target and h is the smoothing parameter [32].
Here are some examples of continuous asymmetric kernels:

- **Beta kernel**: (see Chen [33]).

$$K_{\frac{x}{h}+1,\frac{(1-x)}{h}+1}(u) = \frac{u^{x/h}(1-u)^{\frac{(1-x)}{h}}}{B(\frac{x}{h}+1, \frac{(1-x)}{h}+1)} 1_{\{0 \leq u \leq 1\}}. \quad (7)$$

- **Gamma kernel**: (see Chen [34])

$$K_{\frac{x}{h}+1,h}(u) = \frac{u^{x/h} \exp\{\frac{-u}{h}\}}{h^{\frac{x}{h}+1}\Gamma(x/h+1)}. \quad (8)$$

- **Reciprocal Inverse Gaussian RIG**: (see Scaillet [35])

$$K_{\frac{1}{x-h},\frac{1}{h}}(u) = \frac{1}{\sqrt{2\pi hu}} \exp(\frac{-(x-h)}{2h}(\frac{u}{x-h} - 2 + \frac{x-h}{u})). \quad (9)$$

As already defined above, the Beta kernel is given by the formula eq. 7 (see [33]). It is also called the Beta1 kernel ($K_{\beta_1}(x,h)$).
To reduce the bias near boundaries 0 and 1, Chen proposed a Beta2 estimator defined as follows:

$$\mathrm{K}_{\beta_2}(u) = \begin{cases} K_{\frac{x}{h},\frac{(1-x)}{h}}(u) & \text{si } x \in [2h, 1-2h] \\ K_{\rho_h(x),\frac{(1-x)}{h}}(u) & \text{si } x \in [0, 2h[\\ K_{\frac{x}{h},\rho_h(1-x)}(u) & \text{si } x \in]1-2h, 1] \end{cases} \quad (10)$$

With: $\rho_h(x) = 2h^2 + 2.5 - \sqrt{4h^4 + 6h^2 + 2.25 - x^2 - \frac{x}{h}}$.
We have: $K_{\beta_1}(\alpha,\beta)(0) = K_{\beta_1}(\alpha,\beta)(1)$ which gives an inconsistent estimator.
To solve this problem, Gouriéroux and Montfort [36] have proposed a normalized version of the Beta distribution, which gives the estimators macro-Beta ($Mac-\beta$) and micro-Beta ($Mic-\beta$) respectively :

$$\widehat{f}^{(1)}_{h_n}(x) = \frac{\widehat{f}_{h_n}(x)}{\int_0^1 \widehat{f}_{h_n}(t)dt}, \forall x \in [0,1], \quad (11)$$

$$\widehat{f}^{(2)}_{h_n}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{K_{\beta(x,h_n)(X_i)}}{\int_0^1 K_{\beta(t,h_n)(X_i)dt}} \forall x \in [0,1]. \quad (12)$$

Where the Beta kernel K_β is either K_{β_1} or K_{β_2}.

C. The final kernel modified Champernowne estimator KMCE

The purpose of this section is to show a derivation of the estimator based on the modified Champernowne distribution (KMCE).
The stages of the transformation with the modified Champernowne distribution are as follows: Given a sample $X_i, i = 1,\ldots,N$, we:

- Calculate the parameters $(\hat{\alpha}, \hat{M}, \hat{c})$ of the modified Champernowne distribution as defined above.
- Transform the data set $X_i, i = 1,\ldots,N$, with the transformation function, T:

$$Y_i = T_{\hat{\alpha},\hat{M},\hat{c}}(X_i), i = 1,\ldots,N. \quad (13)$$

The transformation function transforms data into the interval $[0,1]$, and the parameter estimation is used in order to obtain the transformed data as close to a uniform distribution.

- Calculate the classical kernel density estimator on the transformed data $Y_i, i = 1,\ldots,N$:

$$\hat{f}_{trans}(y) = \frac{1}{Nk_y}\sum_{i=1}^{N} K_b(y - Y_i), \quad (14)$$

where $K_b(.) = (1/b)K(.)$, $K(.)$ is the kernel function and k_y is the boundary correction defined as:

$$k_y = \int_{max(-1,-y/b)}^{min(1,(1-y)/b)} K(u)du.$$

- The classical kernel density estimator of the transformed data set results in the KMCE estimator on the transformed scale. Therefore, the estimator of the density of the original data set, $X_i, i = 1,\ldots,N$ is:

$$\widehat{f}(x) = \frac{\widehat{f}_{Trans}(T_{\widehat{\alpha},\widehat{M},\widehat{c}}(x))}{|(T^{-1}_{\widehat{\alpha},\widehat{M},\widehat{c}})'(T_{\widehat{\alpha},\widehat{M},\widehat{c}}(x))|}. \quad (15)$$

The expression of the KMCE is:

$$\widehat{f}(x) = \frac{1}{Nk_{T_{\hat{\alpha},\hat{M},\hat{c}}}(x)}\sum_{i=1}^{N} K_b(T_{\widehat{\alpha},\widehat{M},\widehat{c}}(x) - T_{\widehat{\alpha},\widehat{M},\widehat{c}}(X_i))T'_{\widehat{\alpha},\widehat{M},\widehat{c}}(x). \quad (16)$$

Here are some asymptotic properties (bias and variance) defined in Buch-Larsen et al. [18].
The bias and the variance of $\hat{f}(x)$ are given respectively by:

$$E[\hat{f}(x)] - f(x) = \frac{1}{2}\mu_2(K)b^2((\frac{f(x)}{T'(x)})'\frac{1}{T'(x)})' + o(b^2), \quad (17)$$

$$V[\hat{f}(x)] = \frac{1}{Nb}R(K)T'(x)f(x) + o(\frac{1}{Nb}), \quad (18)$$

as $N \to \infty$, $\mu_2(K) = \int \mu^2 K(u)du$ and $R(k) = \int K^2(u)du$.

III. The semi-parametric approach in the case of the hazard rate function HR

The aim of this section is to propose a new estimator for the hazard rate function in the semi-parametric framework, based on the technique provided in Buch-Larsen et al. [18] using the Champernowne transformation.

A. Form of the new hazard rate function estimator

Using the definition of the hazard rate given by eq. 1, this estimator uses the density and distribution functions f and F respectively.
By replacing f by its estimator (see eq. 16) and F by its estimator (see eq. 20) we obtain the form of the estimator of the hazard rate function which is expressed as follows:

$$\hat{\lambda}(x) = \frac{\frac{1}{Nk_{T_{\hat{\alpha},\hat{M},\hat{c}}}(x)}\sum_{i=1}^{N} K_b(T_{\widehat{\alpha},\widehat{M},\widehat{c}}(x) - T_{\widehat{\alpha},\widehat{M},\widehat{c}}(X_i))T'_{\widehat{\alpha},\widehat{M},\widehat{c}}(x)}{1 - \int_0^x \frac{1}{Nk_{T_{\hat{\alpha},\hat{M},\hat{c}}}(z)}\sum_{i=1}^{N} K_b(T_{\widehat{\alpha},\widehat{M},\widehat{c}}(z) - T_{\widehat{\alpha},\widehat{M},\widehat{c}}(Z_i))T'_{\widehat{\alpha},\widehat{M},\widehat{c}}(z)dz}. \quad (19)$$

Where T is the transformation function and K_b may be replaced by the Beta kernel.

B. Algorithm of the new hazard rate function estimator

The semi-parametric hazard rate function algorithm is divided into two parts: first, the density function then the distribution function are estimated, both using the semi-parametric technique. To elaborate this algorithm, we made appropriate changes to the algorithm proposed by Harfouche and Bareche [37]. The algorithm is as follows:

- Generate a sample $X = X_1, X_2, \ldots X_n$ of size n, from a given distribution.
- Estimate the parameters of the Champernowne generalized distribution (M which is the empirical

median, α and c) as defined in section 2.

- Transform X_i's into Y_i's so that $Y_i = T_{\hat{\alpha},\hat{M},\hat{c}}(X_i)$.
- Estimate the density function of the tranformed data Y_i's according to equation eq. 14.
- Estimate the density function of the original data X_i's according to equation eq. 16.
- Estimate the cumulative function of the original data X_i's according the following equation:

$$\widehat{F}(x) = \int_0^x \frac{1}{Nk_{T_{\hat{\alpha},\hat{M},\hat{c}}}(z)} \sum_{i=1}^{N} K_b(T_{\widehat{\alpha},\widehat{M},\widehat{c}}(z) - T_{\widehat{\alpha},\widehat{M},\widehat{c}}(Z_i)) \quad \text{T}'_{\widehat{\alpha},\widehat{M},\widehat{c}}(z)dz \quad (20)$$

$$= \frac{1}{N} \int_0^x \frac{1}{k_{T_{\hat{\alpha},\hat{M},\hat{c}}}(z)} \sum_{i=1}^{N} K_b(T_{\widehat{\alpha},\widehat{M},\widehat{c}}(z) - T_{\widehat{\alpha},\widehat{M},\widehat{c}}(Z_i)) \quad \text{T}'_{\widehat{\alpha},\widehat{M},\widehat{c}}(z)dz.$$

- Calculate the estimator ratio :

$$\hat{\lambda}(x) = \frac{\hat{f(x)}}{1 - \hat{F(x)}}.$$

IV. Conclusion and perspectives

• The main aim of this article is to use another method that differs from the approaches usually found in insurance and finance (parametric and nonparametric approaches), called the semi-parametric approach, in the context of the hazard rate function.

• After having reviewed the state of the art: for the density function and the hazard rate function in the nonparametric and semi-parametric frameworks, we have proposed a new estimator for the hazard rate function in the semi-parametric sense using the Champernowne transformation and the Beta kernel estimate. We have also elaborated a suitable algorithm for further numerail use.

• As a first perspective, we aim to identify some properties of the new hazard rate function estimator in the semi-parametric framework as: mean, bias and variance. On the other hand, we want to implement under software R the proposed algorithm in order to make simulations (using the Monte Carlo principle) and applications with real data (see Paula et al. [38]) in order to perform comparisons with the works done in the nonparametric framework by Chekkal et al. [21] and in the semi-parametric one by Hua et al. [22].

References

[1] S. Abou-Jouadé, "Sur une condition nécessaire et suffisante de L1-convergence presque complète de l'estimateur de la partition fixe pour une densité," *Comptes rendus de l'Académie des Sciences de Paris, Sér. A-B*, 1976, 283(16), pp 1107-1110.

[2] J. Geffroy, "Sur l'estimation d'une densité dans un espace métrique," *Comptes Rendus de l'Académie des Sciences de Paris, Sér. A-B*, 1976, 278, pp 1449-1452.

[3] N. N. Cencov, "Estimation of an unknown distribution density from observations," *Soviet Math*, 1962, 3, pp 1559-1566.

[4] S. C. Schwartz, "Estimation of probability density by an orthogonal series," *The Annals of Mathematical Statistics*, 1967, pp 1261-1265.

[5] R. Kronmal and M. Tarter, "The estimation of probability densities and cumulatives by Fourier series methods," *Journal of the American Statistical Association*, 1968, 63(323), pp 925-952.

[6] G. Wahba, "Data-based optimal smoothing of orthogonal series density estimates," *The annals of statistics*, 1981, pp 146-156.

[7] D. Bosq, "Estimation suroptimale de la densité par projection," *Canadian Journal of Statistics*, 2005, 31(1), pp 21-37.

[8] N. Saadi and S. Adjabi, " On the estimation of the probability density by trigonometric series," *Communications in Statistics—Theory and Methods*, 2009, 38(19), pp 3583-3595.

[9] G. Wahba, "*Spline models for observational data*," Philadelphia, 1990.

[10] M. Rosenblatt, "Remarks in some nonparametric estimates of a density function," *Annals of Mathematical Statistics*, 1956, 27(3), pp 832-837.

[11] E. Parzen, "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, 1962, 23(3), pp 1065-1076.

[12] M. Hirukawa and M. Sakudo, "Family of the generalised gamma kernels: a generator of asymmetric kernels for nonnegative data," *Journal of Nonparametric Statistics*, 2015, 27(1), pp 41-63.

[13] S. Makhloufi, N. Zougab, Y. Ziane and S. Adjabi, "A family of asymmetric kernels based on log-symmetric distributions," *Communications in Statistics-Simulation and Computation*, 2021, pp 1-18.

[14] C. Marchant, K. Bertin, V. Leiva and H. Saulo, "Generalized Birnbaum–Saunders kernel density estimators and an analysis of financial data," *Computational Statistics and Data Analysis*, 2013, 63, pp 1-15.

[15] C. C. Kokonendji and F. G. Libengué, "Méthode des noyaux associés continus et estimation de densité," *Journées de Statistique de la SFdS 6*, Tunis, 2011.

[16] C. Bolancé, M. Guillen and J. p. Nielsen, "Kernel density estimation of actuarial loss functions," *Insurance: Mathematics and Economics*, 2003, 32(1), pp 19-36.

[17] C. Bolancé, "Optimal inverse Beta(3,3) transformation in kernel density estimation," *Statistics and Operations Research Transactions*, 2010, 34(2), pp 223-238.

[18] T. Buch-Larsen, J.P. Nielsen, M. Guillen and C. Bolancé "Kernel density estimation for heavy-tailed distribution using the Champernowne transformation," *Statistics*, 2005, 6, pp 503-518.

[19] J. Gustafsson, M. Hagmann, J. P. Nielsen and O. Scaillet, "Local transformation kernel density estimation of loss distributions," *Journal of Business and Economic Statistics*, 2009, 27(2), pp 161-175.

[20] U. Balasooriya and C.K. Low, "Modeling insurance claims with extreme observations: Transformed kernel density and generalized lambda distribution," *North American Actuarial Journal*, 2008, 12(2), pp 129-142.

[21] S. Chekkal, K. Lagha and N. Zougab,"Generalized Birnbaum–Saunders kernel for hazard rate function estimation" *Communications in Statistics-Simulation and Computation*, 2023, 52(4), pp 1546-1561.

[22] H. Hua, K, P. N. Patil and D. Bagkavos,"Semiparametric smoothing approach to hazard rate estimation" *Journal of Nonparametric Statistics*, 2017, 29(3), pp 669-693.

[23] D. W. Scott,"Multivariate density estimation: Thory, practise and visualisation" *John willey and sons. Inc*, New York, 1992.

[24] B. W. Silverman,"Density estimation for statistics and data analysis" *Chapman Hall*, London, 1986.

[25] M. Rudemo,"Empirical choice of histograms and kernel density estimators" *Scandinavian Journal of Statistics*, 1982, pp 65-78.

[26] A. W. Bowman,"An alternative mothod of cross-validation for the smoothing of density estimates" *Biometrika*, 1984, 71(2), pp 353-360.

[27] D. W. Scott and G. R. Terrell,"Biased and unbiased cross-validation in density estimation" *Journal of american Statistical association*, 1987, 82(400), pp 1131-1146.

[28] P. Hall, J. Marron and B. U. Park,"Smoothed cross-validation" *Probability theory and related fields*, 1992, 92(1), pp 1-20.

[29] B. U. Park and J. S. Marron,"Comparaison of data-driven bandwidth selectors" *Journal of the American Statistical Assiciation*, 1990, 85(409), pp 66-72.
[30] S. J. Sheather and M. C. Jones,"A reliable data-based bandwith selection method for kernel density estimation" *Journal of the Royal Statistical Society: Series B (Methodological)*, 1991, 53(3), pp 683-690.
[31] N. Zougab, S. Adjabi and C. C. Kokonendji," A bayesian approch to bandwith selection in univariate associate kernel estimation" *Journal of Statistical Theory and Practice*, 2013, 7(1), pp 8-23.
[32] T. Bouezmarni and O. Scaillet," Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data" *Econometric Theory*, 2005, 21(2), pp 390-412.
[33] S. X. Chen," Beta kernels estimators for density functions" *Computational Statistics and Data Analysis*, 1999, 31, pp 131-145.
[34] S. X. Chen," Gamma kernel estimators for density functions" *Annals of the Institute of Statistical Mathematics*, 2000, 52, pp 471-480.
[35] O. Scaillet,"Density estimation using inverse and reciprocal inverse gaussian kernels" *Journal of Nonparametric Statistics*, 2004, 16, pp 217-226.
[36] C. Gourieroux and A. Montfort,"(non) Consistency of the beta kernel estimator for recovery rate distribution" *Working Paper 2006-31, Center for Research in Economics and Statistics*.
[37] Z. Harfouche and A. Bareche,"Semi-parametric approach for approximating the ruin probability of classical risk models with large claims" *Communications in Statistics-Simulation and Computation* , 2021, pp 1-20.
[38] G. Z. Paula, V. Leiva, M. Barros and S. Liu," Robust Statistical modeling using the birnbaum-Saunders-t distribution applied in insurance" *Applied Stochastic Models in Business and Industry*, 2012, 28(1), pp 16-34.

Transmission reliability in WLANs based OFDMA technique

Brahmi Saloua
LaMOS Research Unit, Faculty of Exact Sciences, University of Bejaia, Algeria.
saloua.brahmi@univ-bejaia.dz

Yazid Mohand
LaMOS Research Unit, Faculty of Exact Sciences, University of Bejaia, Algeria.
mohand.yazid@univ-bejaia.dz

Bouhali Abdelhakim
Faculty of Exact Sciences, University of Bejaia, Algeria.
abdelhakim.bouhali@se.univ-bejaia.dz

Bezghiche Micipsa
Faculty of Exact Sciences, University of Bejaia, Algeria.
micipsa.bezghiche@se.univ-bejaia.dz

Abstract—
HE-WLANs (High Efficiency Wireless Local Area Networks) are an evolution of Wi-Fi wireless networks designed to offer enhanced performance, including higher throughput and spectral efficiency. One of the key features of these networks is the use of MU-OFDMA (Multi-user Orthogonal Frequency Division Multiple Access) technology, which allows multiple users to transmit simultaneously on orthogonal subcarriers, by dividing a single 20 MHz transmission channel into numerous narrower transmission sub-channels known as RUs (Resource Units). This work is an improvement of the HMAC protocol with a focus on enhancing transmissions reliability of random access. The simulation results show that the proposal protocol significantly enhances network performance compared to those of HMAC protocol.

Keywords—
High Efficiency WLANs, IEEE 802.11ax Standard, OFDMA Transmission Technique, Multi-User Communications, Random Access, Performance Evaluation and Comparison.

I. INTRODUCTION

The rising prevalence of smart devices and the growing demands of the Internet of Things (IoT) [1] are leading to an increased focus on high-density deployment scenarios in future wireless networks. Locations such as airports, stadiums, shopping malls, and other businesses are set to play a crucial role [2]. This dual importance arises because, on one hand, upcoming wireless networks will require the installation of numerous wireless access points, including base stations (BS) and access points (APs), in limited geographic areas to ensure necessary coverage and capacity. Conversely, future wireless networks must also handle high user densities within a single cell. This includes scenarios like a large number of smartphones in a stadium or a multitude of IoT devices in a smart home or corporate network [3].

Due to the growing utilization of WLANs, the IEEE association has introduced a new working group named 802.11ax, specifically focusing on High Efficiency WLAN. This inventive initiative was initiated back in 2014 with the goal of enabling a substantial user connection, significantly enhancing throughput, and optimizing the utilization of radio resources [4].

Within this novel standard, a revolutionary concept known as RU (Resource Unit) has been incorporated into its PHY (Physical) layer utilizing the OFDMA (Orthogonal Frequency Division Multiple Access) transmission approach. This involves the creation of several distinct non-overlapping RUs with varying widths from a conventional transmission channel [5]. The MAC (Medium Access Control) layer supervises these RUs to facilitate the concurrent service of multiple users, accommodating as many as 9 individuals in what is referred to as MU (Multi-User) access [2].

The AP may schedule DL (Downlink) MU transmissions and UL (uplink) MU transmissions. For DL MU transmissions, since the AP is the initiator of the transmission, user selection and resource distribution between different STAs (stations) do not require any further signaling mechanism. However, for UL MU transmissions, user stations are scheduled by the AP to start their simultaneous transmissions by using a control frame [6].

The access point should have the necessary knowledge of the UL traffic to manage MU communications in the UL direction. Since only the concerned stations are aware of the UL requirements, the access point should initiate a collecting procedure to gather the UL requests. To do this, the IEEE 802.11ax specification enables RUs two access modes: scheduled and random. A deterministic mode that guarantees collision-free access, scheduling access creates a large amount of overhead. In terms of access, the random access mode is fair, but because of collisions, it results in losses [7].

OFDMA is widely applied in cable access networks and wireless communication systems and has been studied for a long time [5]. Many different schedulers have been designed for OFDMA MU access: centralized, dis-

tributed or a hybrid access. This paper focuses on UL OFDMA because DL transmissions are relatively easy to perform in MU OFDMA communications, and in particular we deal with the problem of random access transmissions.

The rest of the paper is organized as follows. In Section II, we describe on the one hand our motivation and on the other hand we review the main existing works about access modes to RUs in Multi-User OFDMA communications. In Sections III, we describe the operation steps of our proposal. In Section IV, we evaluate the proposed protocol. Finally, we conclude the paper in Section V.

II. BACKGROUND

A. PROBLEM STATEMENT

Depending on the supporting technology, multi-user simply implies that transmissions between an AP and numerous clients can occur at the same time. However, when discussing 802.11ax, the MU terminology might be highly confusing. Both OFDMA and MU-MIMO have MU capabilities. In addition to the DLMU-MIMO supported by IEEE 802.11ac, UL-MU-MIMO (which relies on transmitting multiple spatial streams to various stations) is supported by IEEE 802.11ax [8].

802.11ax specifies the usage of both OFDMA and MU-MIMO multi-user technologies. However, OFDMA and MU-MIMO will not be mixed. By subdividing a channel, OFDMA enables multiple user access. By utilising various spatial streams, MU-MIMO enables multiple user access. Using the car and road example from earlier, OFDMA uses a single road divided into multiple lanes for simultaneous use by numerous cars, whereas MU-MIMO uses multiple single-lane roads to reach the same destination [9]. In our work we are interested in MU OFDMA transmission, which allows the implementation of genuine MU communications in the new IEEE 802.11ax standard.

OFDMA is implemented on top of OFDM, with the base station allocating a portion of carriers to each user to allow for multiple simultaneous transmissions. OFDMA employs synchronous medium access, resulting in less congestion (i.e., fewer collisions). As a result, the IEEE 802.11ax Task Group has identified uplink and downlink OFDMA (where the minimum size of a resource unit (RU) is 26 subcarriers) as the essential multi-user feature for improving physical layer efficiency [8].

Unlike DL traffic, which the access point knows very well (since it is local), UL traffic is absolutely unknown to the latter (because it is dispersed over all network stations). As a result, before initiating the UL communication phase, the access point should first collect the UL needs of each station using one of the IEEE 802.11ax collecting modes: scheduled access mode or random access mode [10].

In scheduled access mode, the access point examines all stations to determine whether or not they have UL data to send, allowing each station to broadcast its UL request in a predictable manner. In random access mode, the access point invites only stations with a non-empty queue to communicate their UL demands via a special frame known as the Trigger frame. The stations in question convey their desires simultaneously by selecting a 26-RU at random. Once the access point obtains the necessary information about the stations with UL data, it plans the UL communication phase [7].

Each access mode has benefits as well as drawbacks, thus in order to properly manage the UL request gathering stages, we need a protocol that employs both techniques while minimising their adverse effect.

B. MAIN EXISTING SOLUTIONS

Deng et al. [11] proposes the adoption of the GRAP (Group Randomly Addressed Polling) protocol, as originally introduced by [12], for the purpose of gathering uplink requirements from network stations engaged in OFDMA MU communications. This protocol aims to facilitate the collection of UL needs by inviting network STAs to communicate their requirements across several stages, utilizing the conventional Beacon control frame. The Beacon frame not only specifies the number of stages involved but also delineates the structure of the transmission channel. During each stage, STAs with data in their queues randomly select a transmission subchannel through which they transmit their UL requests. The access point (AP) subsequently determines the stage with fewer collisions and prioritizes it. Even though the GRAP protocol improves the success rate of UL requests, it does introduce a significant overhead.

Lee et al. [13] merged both random and scheduling access modes into RUs to formulate the HMAC (Hybrid Multiple Access Coordination) MAC protocol. The primary goal was to minimize the overhead arising from the collection of uplink requirements of STAs. For this purpose, STAs designated to receive DL traffic utilize allocated RUs to transmit their UL requests in scheduling mode. Meanwhile, remaining STAs utilize unallocated RUs to send their UL requests in random mode. While this approach effectively manages the collection time for UL requests, an issue arises when STAs solely have UL data streams to receive from the AP; they face limitations in sending UL requests if all RUs are assigned for DL transmission.

Lin et al. [14] developed a MAC protocol known as G-OFDMA (Group based OFDMA) to connect a

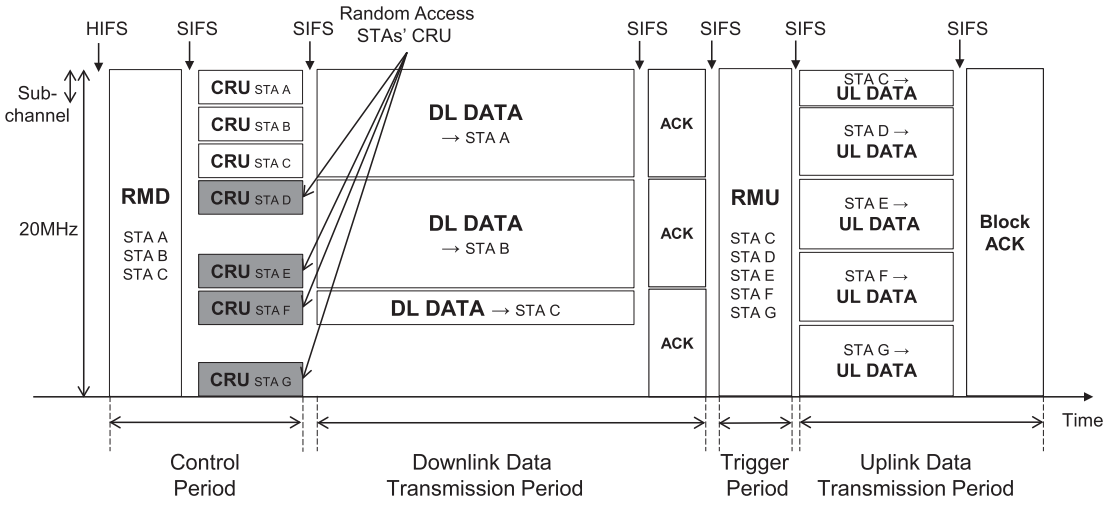


Figure 1. HMAC protocol.

larger number of users in high density WLANs. To accomplish this, the AP invites STAs to distribute their UL requests in random mode throughout a time interval comprised of numerous time slots using the traditional Beacon frame. Each STA with a non-empty queue then chooses a transmission subchannel and a time window at random to deliver its UL request using the traditional RTS frame. This protocol allows multiple STAs to have the opportunity to send their requests over an extended contention period, but, on the other hand, it generates considerable overhead, and the other STAs must wait until the end of a long series of data transmissions before trying again.

Zhou et al. [15] introduced a MAC protocol named DRA-OFDMA (Double Random Access-based OFDMA), which operates within the framework of the 802.11ax standard control frames (specifically BSR and Trigger). To achieve this, the STAs access the RUs and distribute their UL requests in a random manner. Each STA with a successful BSR is given one 26-RU by the AP to transmit its UL stream in scheduling mode. If any remaining 26-RUs are not allocated, they will be sent to STAs with failed BSR so they can transmit their video streams in random mode. Even though the DRA-OFDMA protocol integrates the concept of service differentiation, it doesn't guarantee the successful transmission of video streams.

Avdotin et al. [16] developed a resource allocation algorithm CRA (Cyclic Resource Assignment) aimed at minimization of RTA (Real Time Applications) data transmission delay in Wi-Fi networks-based OFDMA transmission technique. It uses the two access modes in the same slot time to assign RUs to the UL STAs. The CRA method allows the AP to keep track of which RUs have experienced collisions. When there are no collisions in a slot, only f RUs are assigned for RA in the following slot (the remaining RUs are reserved for non-RTA transmissions). In this case, the AP knows that there are no STAs in the network that have failed to transmit critical data. Therefore, only those STAs that have recently received packets are able to broadcast in these f RUs. For optimisation of resource allocation for STA RTAs, the CRA algorithm is based on idealised assumptions such as only 26 RUs, saturate traffic, fixed data size, etc.

As observed, limited research has been dedicated to investigating access modes within OFDMA Multi-User communications. It's crucial to emphasize the significance of access modes in the context of OFDMA transmission. These modes facilitate the transmission of UL requests from stations to the access point.

III. PROPOSAL

In the operation role of the HMAC protocol (see Fig. 1), the data transmission phase in the UL communication depends on the "Control Period". In other words, the number of stations participating in the UL phase includes stations that have UL data to transmit among the stations participating in the DL phase, plus the stations that have sent their CRUs (Clear-to-Receive-with-UL-request) in random access mode. So, after collecting the CRUs from stations in centralized access scheduling, the other stations that wish to transmit UL data will contend on the remaining unallocated sub-channels. Among these stations, there will be some that collide when two or more stations choose the same RUs, and there will be those that succeed in obtaining RUs and sending their CRUs.

To reduce the collision rate and improve UL service rate by increasing the number of UL stations successfully

send there requests, as a new improvement of HMAC, we involves separating the two phases of scheduling and random access modes and executing them sequentially. The goal is to allow competition among stations in random access only on the free sub-channels but reserve all the sub-channels of the channel for nine (9) RUs in a 20 MHz channel.

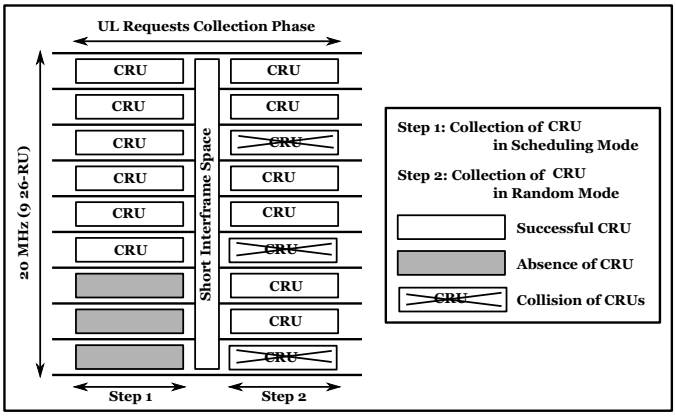


Figure 2. UL request collection steps.

The new control period or the signaling phase is shown in Fig. 2, the UL request collection step is made up of 2 steps: the first step is reserved for scheduling access which means that the stations programmed for reception of the DL streams are invited to send their request by sending a CRU control frame on a 26-RU which was reserved for him. By contrary, the rest of the stations with a non-empty queue are invited to send their request by entering into competition in the second collection stage by choosing a 26-RU randomly.

IV. PERFORMANCE EVALUATION

In this section, we assess and contrast the performance of the proposed protocol to that of the HMAC protocol. We analyse the following performance metrics to do this:

- **Collision Ratio:** it represent the average percentage of UL requests transmitted in random mode per cycle.
- **UL Service Rate of STAs in RA:** it represents the average number of STAs which have successfully sent their request in random access mode and served it in the UL direction per communication cycle.
- **UL throughput of STAs in RA:** it represents the average amount of UL data sent by STAs which have successfully sent their request in random access mode per communication cycle.

To obtain the aforementioned measurements, we implemented the proposal and HMAC protocols' operating rules using the Matlab programming language. The most important PHY and MAC parameters that we have put forward to evaluate the performance are defined in the Tab. I.

Table I. Simulation parameters.

Parameter	Value
Channel Bandwidth	20 MHz
Type of RU	26, 52, 106, 242
Data Packet PHY rates (Mb/s)	11.8, 23.5, 50, 114.7
Control Packet PHY rates (Mb/s)	0.9, 1.8, 3.8, 8.6
Short Inter-Frame Space (SIFS)	16 μs
Gard interval	0.8 μs

In Fig. 3, we evaluate the Collision ratio of the proposed protocol and compare it with HMAC protocol according to the number of stations in the network. We show that the collision ratio of UL requests in both protocols is increasing when the number of stations increases in the network and it reaches 100% after 40 STAs, because the two protocols use random access mode to allow stations with non-empty queues to send their UL needs. However, we note that with the proposed protocol, this increase is less severe since it uses the large width of the channel in random access mode. In fact, just a portion of the channel is used for random access with the HMAC protocol.

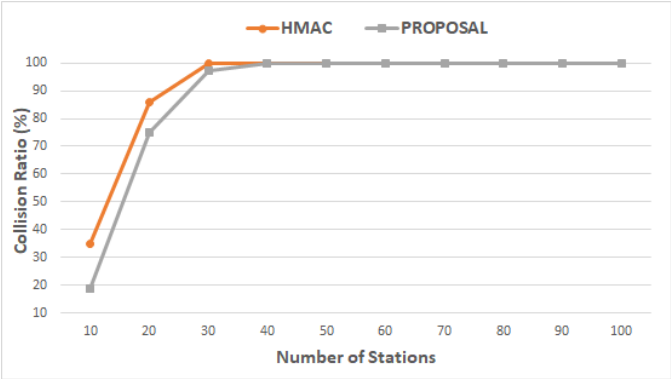


Figure 3. Collision ratio versus number of stations.

In Fig. 4, we evaluate the UL service rate of STAs in RA of the proposed protocol and compare it with HMAC protocol according to the number of stations in the network. We observe that, the greater the number of stations in the network, the smaller the UL service rate. This is due to the increase in the collision rate of requests sent by STAs in random mode. Indeed, our protocol exceeds HMAC by approximately 74% this due to the improvement of the collision ratio of the proposed protocol.

In Fig. 5, we study the UL throughput of STAs in RA of the proposed protocol and compare it with HMAC protocol according to the number of stations in the network. We notice that the evolution of the UL service rate follows the evolution of the UL service rate in the both protocols. due to the fact that UL throughput depends directly on the number of UL service STAs which in turn depends on the collision ratio.

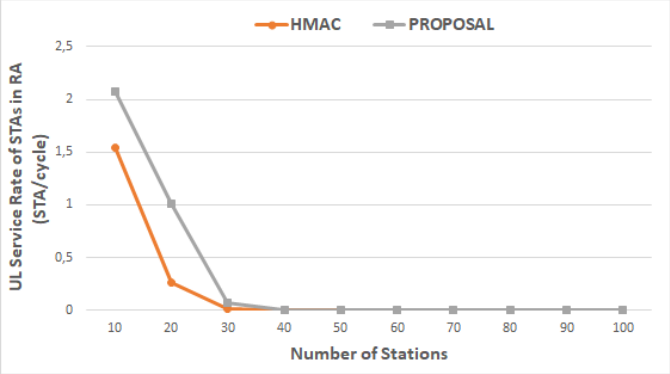


Figure 4. UL service rate of STAs in RA versus number of stations.

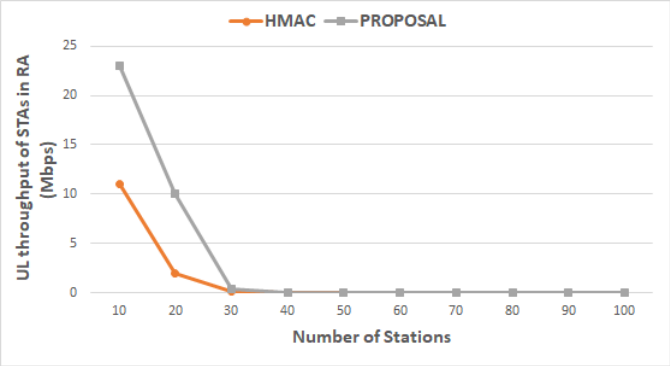


Figure 5. UL throughput of STAs in RA versus number of stations.

V. CONCLUSION

In this research, we concentrate our investigation on random access mode to resource units in OFDMA Multi-User communications based on IEEE 802.11ax WLAN networks. Indeed, through this access mode, network stations can transmit their UL traffic requirements to the access point, allowing for potential scheduling during the UL communication phase. In this work, we improved the HMAC protocol, in particular we have improved the UL request collection part. We have retained the first request collection step for scheduled access. Additionally, we have introduced a second step for random access, which spans the entire bandwidth of the transmission channel, this serves the dual purpose of reducing collision ratio and increasing the UL service rate. The performance evaluation demonstrates that the suggested HMAC protocol modification greatly enhances network performance in terms of collision rate, service rate, and throughput in UL communications.

REFERENCES

[1] M. Yang, L. Bo, and Y. Zhongjiang, "MAC Technology of IEEE 802.11 ax: Progress and Tutorial," *Mobile Networks and Applications*, 2021, 26, pp. 1122-1136.

[2] S. Brahmi, and M. Yazid, "Towards a Fair Allocation and Effective Utilization of Resource Units in Multi-User WLANs-based OFDMA technology," *Computer Networks*, 2023, 224, p. 109639.

[3] S. Avallone, P. Imputato, G. Redieteab, C. Ghosh, and S. Roy, "Will OFDMA improve the performance of 802.11 WiFi networks?," *IEEE Wireless Communications*, 2021, 28(3), pp. 100-107.

[4] S. Brahmi, M. Yazid, and M. Omar, "Multiuser Access via OFDMA Technology in High Density IEEE 802.11 ax WLANs: A Survey," *In: 2020 Second International Conference on Embedded & Distributed Systems (EDiS)*, 2020, pp. 105–110, IEEE.

[5] S. Tutelian, D. Bankov, D. Shmelkin, and E. Khorov, "Ieee 802.11 ax ofdma resource allocation with frequency-selective fading," *Sensors*, 2021, 21(18), p. 6099.

[6] B. Bellalta, and K. Kosek-Szott, "AP-initiated multi-user transmissions in IEEE 802.11 ax WLANs," *Ad Hoc Networks*, 2019, 85, pp. 145-159.

[7] S. Brahmi, M. Yazid, "Towards a Hybrid Access to Resource Units in Multi-user OFDMA Communications," *In: International Conference on Artificial Intelligence in Renewable Energetic Systems*, 2021, pp. 421–429, Springer.

[8] M. S. Afaqui, E. Garcia-Villegas, and E. Lopez-Aguilera, "IEEE 802.11 ax: Challenges and requirements for future high efficiency WiFi," *IEEE wireless communications*, 2016, 24(3), pp. 130–137.

[9] D. D. Coleman and D. A. Westcott, "802.11ax: High Efficiency (HE)," *5th ed. John Wiley & Sons*, 2018, Chap. 19, pp. 784–814.

[10] R. Ali, S. W. Kim, B. S. Kim and Y. Park, "Design of MAC layer resource allocation schemes for IEEE 802.11 ax: Future directions", *IETE Technical Review*, 2018, 35(1), pp. 28-52.

[11] D. J. Deng, K. C. Chen, and R. S. Cheng, "IEEE 802.11 ax: Next generation wireless local area networks," *In 10Th international conference on heterogeneous networking for quality, reliability, security and robustness*, 2014, pp. 77-82, IEEE.

[12] K. C. Chen, "Medium access control of wireless LANs for mobile computing," *IEEE Network*, 1994, 8(5), pp. 50-63.

[13] J. Lee, and C. Kim, "An efficient multiple access coordination scheme for OFDMA WLAN," *IEEE Communications Letters*, 2016, 21(3), pp. 596-599.

[14] S. Lin, H. Qi, X. Wen, Z. Lu, and Z. Hu, "An efficient group-based OFDMA MAC protocol for multiuser access in dense WLAN systems," *in: IEEE International Conference on Communications Workshops*, 2018, pp. 1–6.

[15] R. Zhou, B. Li, M. Yang, Z. Yan, and A. Yang, "DRA-OFDMA: Double random access based QoS oriented OFDMA MAC protocol for the next generation WLAN," *Mobile Networks and Applications*, 2019, 24, pp. 1425-1436.

[16] E. Avdotin, D. Bankov, E. Khorov, and A. Lyakhov, "Enabling massive real-time applications in IEEE 802.11 be networks", *In 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2019, pp. 1–6, IEEE.

Optimisation de la disponibilité des engins au sein de l'entreprise portuaire de Skikda

Cherfaoui Bachir
unité de recherche LaMOS ,département technologie, faculté de technologie, université de Bejaia
bachir.cherfaoui@univ-bejaia.dz

Hamidcha Mossaab
déparetement RO , faculté des sciences exactes, université de Bejaia
mosaab.hamidcha@se.univ-bejaia.dz

Aissani Djamil
unité de recherche LaMOS, faculté des sciences exactes, université de Bejaia
djamil.aissani@univ-bejaia.dz

***Résumé*—**
Les systèmes de manutention dans les entreprises portuaires jouent un rôle crucial, ce qui fait de leur disponibilité un atout de première importance. Dans ce travail on s'intéresse à l'optimisation des instants et le nombre de remplacements préventifs imparfaits pour maximiser la disponibilité de système de manutention de port de skikda.

Un modèle de maintenance imparfaite basé sur le processus de quasi-renouvèlement est appliqué sur quelques engins dans le but d'optimiser le nombre de remplacement préventifs et les instants de remplacement préventifs. Des résultats numériques ont été présentés afin de monter l'efficacité du modèle en question.

Mots-Clés—
disponibilité, maintenance imparfaite, quasi-renouvèlement

I. INTRODUCTION

Dans l'univers des entreprises portuaires, les systèmes de manutention jouent un rôle central, étant un maillon essentiel dans la chaîne logistique de ces installations. La disponibilité de ces systèmes est un facteur critique pour garantir un fonctionnement fluide, efficace et rentable des opérations portuaires.

Afin d'assurer une disponibilité maximale des systèmes nous devrons appliquer un plan maintenance efficace. Plusieurs travaux ont été réalisés dans ce domaine [1-3]. Entre autres, on peut citer le travail de Bouhamou F et Oukaour, N. [3], qui ont mené une étude d'analyse du retour d'expérience pour la planification de la maintenance du parc des chariots élévateurs au Port de Bejaïa.Une autre étude a été réalisée au sein de l'entreprise MAC-SOUMAM par Chahboune S et Hammou R [4], l'objectif de cette étude était d'analyser la disponibilité des équipements de la chaîne de montage B de l'entreprise MAC-SOUMAM, en se basant sur le calcul de la fiabilité et de la maintenabilité des équipements étudié. Arturo [6] a proposé un modèle de maintenance imparfaite dans le but d'optimiser la disponibilité des systèmes en dégradation.

Dans notre étude, nous allons nous intéresser à l'application d'une politique de maintenance imparfaite basée sur le processus de quasi-renouvèlement pour optimiser la disponibilité des engins de manutention d'une entreprise portuaire, en prenant en considération deux facteurs essentiels, le facteur de la réduction de la durée moyenne de vie du système et le facteur l'augmentation de la durée de réparation du système.
Le reste de cet article s'organise comme suit, la section 2 est dédiée à la présentation mathématique du modèle de maintenance appliqué. La section 3 consiste en une application numérique pour rechercher les paramètres optimaux de la politique en question, et une étude de sensibilité est effectuée.

II. MODELE MATHEMATIQUE

Notations
$A(T,k)$: Disponibilité moyenne.
$F_1(t)$: Fonction de distribution.
$R_1(t)$: Fonction de fiabilité.
μ : Durée de vie moyenne de système.

η : Durée de réparation moyenne.

α : Facteur de réduction de l'âge du système $0 \prec \alpha \prec 1$

β : Facteur d'augmentation des durées de réparation ($\beta \succ 1$)

λ : Paramètre d'échelle.

θ : Paramètre de forme.

T : L'âge fixe dans lequel une unité est soumise à la maintenance préventive $T \succ 0$.

k : Nombre de réparations imparfaites avant d'effectuer la remise à neuf du système.

La politique de maintenance imparfaite basée sur le processus de quasi-renouvèlement est schématisée dans la figure fig .1. Son modèle peut être décrit de la manière suivante :

On a

x_1 : Durée de vie pour une nouvelle unité.

y_1 : Première réparation imparfaite.

αx_1 : Temps de vie de l'unité après la première réparation imparfaite.

βy_1 : Temps de réparation imparfaite.

αx : Temps de vie moyen.

βy : Temps de réparation moyen.

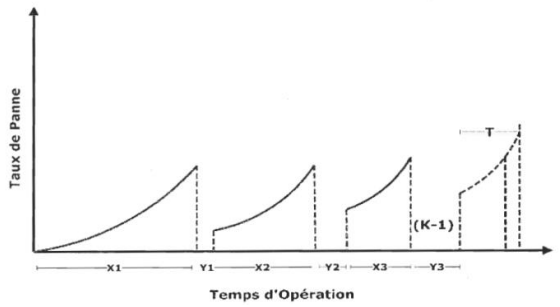


Figure 1. Schéma du modèle quasi-renouvellement.

Le temps de réparation augmente avec l'augmentation de nombre de réparations avec une valeur β > 1 cela veut dire que la durée de vie de l'unité diminue de α (0 < α <1) après chaque réparations imparfaites réalisées.

On suppose qu'à l'instant $t = 0$, on commence l'observation de système. Si après x_1 unités de temps, le système tombe en panne et que le temps pour effectuer une réparation imparfaite est y_1 alors, selon le processus de quasi-renouvellement, la prochaine réparation aura lieu après x_2 unités de temps et la durée de réparation sera y_2 tel que:

$$x_2 = \alpha x_1 \quad \text{avec } 0 < \alpha < 1 \tag{1}$$

$$y_2 = \beta y_1 \quad \text{avec } \beta > 1 \tag{2}$$

à La (k-1) ème réparation en aura :

$$x_{k-1} = \alpha^{k-2} x_1 \tag{3}$$

tel que :

$$x_2 = \alpha x_1, x_3 = \alpha^2 x_1, x_4 = \alpha^3 x_1, ..., x_{k-1} = \alpha^{k-2} x_1, x_k = \alpha^{k-1} x_1$$

De même que pour y_{k-1} $y_{k-1} = \beta^{k-2} y_1$ tel que

$$y_2 = \beta y_1, y_3 = \beta^2 y_1, y_4 = \beta^3 y_1, ..., y_{k-1} = \beta^{k-2} y_1, y_k = \beta^{k-1} y_1$$

La fonction de disponibilité est le rapport entre le temps de bon fonctionnement et le temps moyen total donnée comme suit :

$$A(T,k) = \frac{U(T,k)}{D(T,k)} \tag{4}$$

Avec

$U(T,k)$: Le temps moyen de fonctionnement qui s'écrit comme suit

$$U(T,k) = \frac{\mu(1-\alpha^{k-1})}{1-\alpha} + \int_0^T R_1(\frac{1}{\alpha^{k-1}} t)dt \tag{5}$$

Et $D(T,k)$: temps moyen total donné par la formule

$$D(T,k) = \frac{\mu(1-\alpha^{k-1})}{1-\alpha} + \frac{\eta(\beta^{k-1}-1)}{\beta-1} + \int_0^T R_1(\frac{1}{\alpha^{k-1}} x)dx \tag{6}$$

Donc la disponibilité s'ecrit comme suit :

$$A(T,k) = \frac{\frac{\mu(1-\alpha^{k-1})}{1-\alpha} + \int_0^T R_1(\frac{1}{\alpha^{k-1}} t)dt}{\frac{\mu(1-\alpha^{k-1})}{1-\alpha} + \frac{\eta(\beta^{k-1}-1)}{\beta-1} + \int_0^T R_1(\frac{1}{\alpha^{k-1}} x)dx} \tag{7}$$

III. RESULTATS ET DISCUSSION

Dans la première partie de cette section nous cherchons à trouver le couple optimale (k , T) pour trois engins au sein de l'entreprise portuaire en appliquant la stratégie de remplacement préventif imparfait définit dans la section précédente.

A l'aide de données collectées à l'entreprise nous avons calculé les différents paramètres à utiliser dans le modèle en question (voir Tab 1).

Tableau 1. Les différents paramètres liés au trois engins.

	α	β	η	μ	λ	θ
Engin 1	0.95	1.05	21	182.18	200.14	1.41
Engin 2	0.95	1.05	14	142.85	143.66	1.14
Engin 3	0.95	1.05	26	91.40	78.30	0.77

Nous avons implémenté un programme sur Matlab qui nous permet de maximiser la disponibilité A , en optimisant le nombre de remplacement préventifs k et la périodicité de remplacement T . Les résultats sont donnés dans le tableau suivant (Tab. 2).

Tableau 2. Valeurs optimales de k , T et A

	k^*	T^*	A^*	A
Engin1	4	3559	0.96	0.87
Engin2	2	3789	0.95	0.90
Engin3	2	6116	0.89	0.81

D'après Tab.2 on voit une amélioration considérable concernant la disponibilité après application de la stratégie de maintenance imparfaite.

Pour évaluer l'impact des variables α (facteur de réduction de l'âge du système), β (facteur d'augmentation des durées de réparation) et k (nombre de réparations imparfaites) sur les optimal trouvés T^* et A^*, nous procéderons à une étude de sensibilité à ce sujet.
Dans Tab. 3-8 suivants, nous avons présenté les variations de la disponibilité et du temps optimal T en fonction des facteurs α et β, ainsi que du nombre de réparations imparfaites K. pour les 3 engins en question.

Tableau 3. Variation de la disponibilité en fonction de β (Engin 1).

K	β	T^*	A^*
	1.1	569.5566	0.9669
	1.3	569.5566	0.9669
2	1.6	569.5566	0.9669
	1.8	569.5566	0.9669
	2.0	569.5566	0.9669
	1.1	483.7742	0.9360
	1.3	486.0576	0.9164
5	1.6	493.9124	0.8799
	1.8	549.1198	0.8511
	2.0	502.5927	0.8189
	1.1	422.2789	0.9141
	1.3	434.3975	0.8517
8	1.6	442.9647	0.7010
	1.8	445.3906	0.5729
	2.0	446.6446	0.4429

Tableau 4. Variation de la disponibilité en fonction de α (Engin 1).

k	α	T^*	A^*
	0.2	134.8201	0.9474
	0.4	250.7629	0.9545
2	0.6	353.1622	0.9599
	0.8	496.5000	0.9642
	0.9	545.7490	0.9660
	0.2	1.9900	0.8131
	0.4	18.8850	0.8517
5	0.6	91.2833	0.8892
	0.8	260.986	0.9212
	0.9	396.1107	0.9344
	0.2	1.9900	0.6973
	0.4	1.9900	0.7542
8	0.6	19.8538	0.8191
	0.8	130.2835	0.8846
	0.9	299.5278	0.9130

Tableau 5. Variation de la disponibilité en fonction de β (Engin 2).

K	β	T*	A*
	1.1	599.8367	0.9517
	1.3	599.1200	0.9510
2	1.6	597.8557	0.9455
	1.8	596.4337	0.9441
	2.0	595.2561	0.9420
	1.1	598.9089	0.9081
	1.3	599.6758	0.8812
5	1.6	599.7626	0.8321
	1.8	599.7880	0.7944
	2.0	599.7101	0.7536
	1.1	517.8364	0.8780
	1.3	544.0400	0.7952
8	1.6	566.0670	0.6132
	1.8	570.1074	0.4756
	2.0	565.7361	0.3497

Tableau 6. Variation de la disponibilité en fonction de α (Engin 2).

k	β	T*	A*
	1.1	599.9989	0.8720
	1.3	599.9989	0.8720
2	1.6	599.9989	0.8720
	1.8	599.9989	0.8720
	2.0	599.9989	0.8720
	1.1	599.7817	0.7739
	1.3	599.8113	0.7197
5	1.6	599.9987	0.6318
	1.8	599.9988	0.5723
	2.0	599.9989	0.5143
	1.1	522.3240	0.7137
	1.3	533.7738	0.5736
8	1.6	533.4689	0.3544
	1.8	524.4106	0.2390
	2.0	515.0022	0.1570

Tableau 7. Variation de la disponibilité en fonction de β (Engin 3).

k	α	T*	A*
	0.2	195.0645	0.8083
	0.4	398.6028	0.8311
2	0.6	599.7447	0.8490
	0.8	599.8311	0.8632
	0. 9	599.8886	0.8692
	0.2	1.9900	0.5048
	0.4	29.1253	0.5736
5	0. 6	131.9086	0.6528
	0.8	398.4068	0.7327
	0.9	586.3415	0.7695
	0.2	1.9900	0.3505
	0.4	1.9900	0.4183
8	0.6	27.9276	0.5149
	0.8	191.2103	0.6424
	0.9	380.2809	0.7108

Tableau 8. Variation de la disponibilité en fonction de β (Engin 3).

k	α	T*	A*
	0.2	195.0645	0.8083
	0.4	398.6028	0.8311
2	0.6	599.7447	0.8490
	0.8	599.8311	0.8632
	0. 9	599.8886	0.8692
	0.2	1.9900	0.5048
	0.4	29.1253	0.5736
5	0. 6	131.9086	0.6528
	0.8	398.4068	0.7327
	0.9	586.3415	0.7695
	0.2	1.9900	0.3505
	0.4	1.9900	0.4183
8	0.6	27.9276	0.5149
	0.8	191.2103	0.6424
	0.9	380.2809	0.7108

L'analyse des résultats obtenus permet de tirer plusieurs interprétations concernant l'optimisation de la disponibilité des engins au sein de l'entreprise portuaire de Skikda.

a) Impact de la variable α (Facteur de réduction de l'âge du système) :

On remarque pour les 3 engins que l'augmentation de α conduit à une augmentation de la périodicité optimale de maintenance préventive T_{opt} et de la disponibilité maximale $A_{\max}$. Cela indique qu'augmenter l'âge du système a un impact positif sur la disponibilité des engins. Le facteur α dans le modèle de maintenance imparfaite quasi-renouvellement représente le facteur de réduction de l'âge du système (l'engin dans notre cas). Ainsi, une augmentation de ce facteur implique une diminution de la dégradation de la durée de vie moyenne de l'engin, cela signifie une dégradation minimale.

b) Impact de la variable β (Facteur d'augmentation des durées de réparation) :

Pour les trois engins on constate que l'augmentation de β n'a pas d'impact significatif sur les valeurs optimales de disponibilité T_{opt} et de disponibilité maximale $A_{\max}$. Cela suggère que l'augmentation des durées de réparation n'a pas d'effet notable sur la disponibilité des engins dans ce contexte.

c) Variations des valeurs optimales avec k

Pour les 3 engins, on observe que l'augmentation de k conduit à une diminution de la disponibilité maximale. Cela signifie qu'un nombre plus élevé de réparations imparfaites entraîne une mauvaise disponibilité globale.

Les résultats indiquent que les variables α et K ont un impact significatif sur les valeurs optimales de T_{opt} et $A_{\max}$. Une diminution de α et une augmentation de K entraînent une diminution de T_{opt} et $A_{\max}$, ce qui signifie que des niveaux plus bas de réparations (en terme de qualité de réparation) et des nombres plus élevés des réparations imparfaites peuvent conduire à une mauvaise disponibilité.

IV. CONCLUSION

Dans ce travail nous nous sommes intéressés à l'application d'une stratégie de maintenance imparfaite basée sur le processus de quasi-renouvellement sur des engins de manutention de port de skikda dans le but d'optimiser la disponibilité de ces derniers.

Les résultats obtenus montrent que l'augmentation de la durée de réparation avec une efficacité de la maintenance réduite diminue la disponibilité des engins. Lorsque l'efficacité de la maintenance est élevée, on peut atteindre une disponibilité maximale, même en cas de longues durées de réparation. Par conséquent, il est judicieux de conclure que l'accent devrait être mis sur l'amélioration de l'efficacité de la maintenance imparfaite plutôt que d'augmenter les temps de réparation. De plus on constate qu'il est préférable de diminuer le nombre de réparations imparfaite k lorsque les durées moyennes des interventions sont grandes β >>.

REFERENCES

[1] K. T. Huynh and Antoine Grall. Modèle de maintenance conditionnelle imparfaite avec mémoire pour des systèmes à dégradation continue. Congrès Lambda Mu 21 " Maîtrise des risques et transformation numérique : opportunités et menaces ", Reims, France, Oct 2018.

[2] H. Lala, Modélisation et optimisation de la maintenance préventive des équipements de production pétroliers en Algérie, thèse doctorat, université de Constantine, 2022.

[3] T. Nakagawa, Optimum policies when preventive maintenance is imperfect. *IEEE Transactions on Reliability*, vol. 28, no. 4, pp. 331–332, 1979.

[4] N. Oukaour and F. Bouhamou. Analyse du retour d'expérience pour la planification de la maintenance du parc des chariots élévateurs au port de Béjaïa. Mémoire d'ingéniorat, Département de Recherche Opérationnelle, Université de Béjaia, 1998.

[5] R. Hammou and S. Chahboune. Étude de la fiabilité et de la disponibilité des équipements au niveau de l'entreprise mac-soum, Mémoire de master, Département de Recherche Opérationnelle, Université de Béjaia , 2016.

[6] M. Arturo, Optimisation de la disponibilité des systèmes assujettis à la maintenance imparfaite. PhD thèses, Université Laval, 2008.

Wireless Network Simulation: A Practical Case Study on 802.11be

Goutal Abdelhak
Laboratory LAMOS, University of Bejaia
abdelhak.goutal@univ-bejaia.dz

Bouallouche-Medjkoune Louiza
Laboratory LAMOS, University of Bejaia
louiza.medjkoune@univ-bejaia.dz

Moktefi Mohand
Laboratory LAMOS, University of Bejaia
mohand.moktefi@univ-bejaia.dz

***Abstract*—**
The increasing demand for high-performance wireless networks has necessitated the development of advanced simulation techniques to optimize their design and performance. In this study, we present a practical case study focused on the simulation of a wireless network using the 802.11be EHT (Enhanced High Throughput) standard. Through extensive simulations, we explore the capabilities and performance of the network, with particular emphasis on throughput and latency. Our study showcases the effectiveness of simulation in evaluating and optimizing wireless networks, offering valuable insights for network designers and operators. The findings from this case study provide valuable guidance for leveraging the potential of the 802.11be standard in real-world wireless deployments.

Keywords—
Network simulation, Simulation Model, Wireless Communication, Channel Modeling, 802.11be.

I. INTRODUCTION

Wireless networks play a crucial role in our daily lives by providing ubiquitous connectivity for a wide range of applications, from personal communications to enterprise systems. They enable us to stay connected, access information, share data, and communicate with others, transforming the way we work, learn, and entertain ourselves. However, with the exponential growth of connected devices and evolving requirements in terms of quality of service (QoS), throughput, latency and security, it is crucial to continuously improve the performance of wireless networks to meet the increasing demands of users.

Wireless network simulation offers a powerful and indispensable approach to evaluate and enhance the performance of these networks. By using sophisticated simulation models, it allows for the faithful reproduction of the behavior of wireless networks and the study of their performance in controlled and repeatable environments. Simulation provides a flexible experimental framework to test new configurations, evaluate protocols, optimize strategies, and anticipate potential issues. Moreover, simulation enables these experiments to be conducted at a lower cost and within significantly shorter timeframes compared to real-world deployments, making it an essential tool for researchers, engineers, and decision-makers.

In this article, we will emphasize the importance of simulation in evaluating and improving wireless network performance. We will explore how simulation enables the faithful replication of essential characteristics of wireless networks, such as interferences, radio wave propagation conditions, traffic variations, and user behaviors, for study in realistic scenarios. We will discuss the advantages and limitations of simulation, highlighting its relevance for research, development, and optimization of wireless networks.

In the following section, we will review the state-of-the-art of existing simulators for wireless networks, showcasing their categories, features, and functionalities. Subsequently, we will present our specific simulator model used to evaluate the performance of 802.11be in high-density scenarios. We will provide a detailed description of the features and functionalities of our simulator. Next, we will delve into the simulation part of 802.11be, where we will present the simulation objectives, the methodology employed, and the results obtained in terms of throughput and latency. Finally, we will discuss the simulation results, analyze the performance of 802.11be, explore the factors influencing these performances, and propose perspectives for enhancing wireless network performance.

II. STATE OF THE ART OF EXISTING SIMULATORS FOR WIRELESS NETWORKS

A. *Discrete Event-based Simulators*

Discrete event-based simulators are among the most commonly used in wireless network simulation. They model the behavior of network nodes, communication channels, and protocols as discrete events, where each event triggers an action in the system. These simulators allow for detailed modeling of events and actions in the network. They provide the ability to specify specific behaviors for each network node, such as packet

sending and receiving, channel switching, collisions, etc. This allows for a detailed study of interactions between nodes and protocols. Discrete event-based simulators use discrete time management, where time is divided into discrete intervals Fig. 1. Each event is scheduled and executed at a specific time in the simulated time. This enables accurate measurement of network performance and synchronization of actions between nodes. These simulators are often accompanied by specific libraries and modules that provide additional functionality. For example, libraries for mobility models, signal propagation models, routing protocols, etc. This facilitates the addition of features and extension of the simulator to meet specific simulation needs.

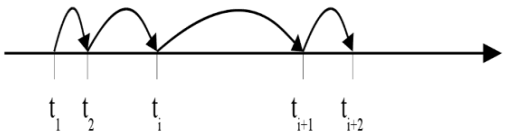


Figure 1. Discrete Event-based Simulators [9]

1) Examples of discrete event-based simulators

a) NS-3 (Network Simulator 3)

Is a highly popular open-source network simulator based on discrete events *[1]*. It provides a wide range of models and protocols to simulate wireless networks, including Wi-Fi, ad hoc networks, cellular networks, etc. NS-3 offers great flexibility and high modularity, making it a common choice for research and development in the field of wireless networks.

b) OMNeT++

Is another discrete event-based simulator widely used in modeling and simulation of wireless networks. It offers an extensible and modular simulation platform, with libraries and modules dedicated to wireless networks. OMNeT++ supports detailed modeling of nodes, channels, and protocols, allowing for accurate and realistic simulation *[2]*.

c) OPNET (Optimized Network Engineering Tool)

Is a commercial discrete event-based simulator widely used in industry for simulating wireless networks. It offers advanced features for protocol modeling, quality of service management, performance analysis, etc. OPNET is known for its ability to simulate large-scale wireless networks and for its detailed performance analysis tools *[3]*.

2) Advantages

a) *Flexibility*

Discrete event-based simulators offer great flexibility in modeling the behavior of network nodes and protocols. They allow for the definition of specific events, such as packet sending or receiving, channel switching, etc., enabling detailed simulation of interactions between different network elements.

b) *Time control*

These simulators allow for control of simulation time, facilitating accurate performance measurement and comparison between different configurations.

c) *Extensibility*

Discrete event-based simulators are often accompanied by libraries and modules that enable the extension of their functionality. This facilitates the addition of new protocols, mobility models, channel models, etc., according to the simulation needs.

3) Disadvantages

a) *Complexity*

Modeling discrete events can be complex, especially when simulating wireless networks with dynamic behaviors. Designing and implementing simulation models may require deep expertise and a detailed understanding of protocols and network interactions.

b) *Simulation time*

Discrete event-based simulators may require significant computational resources and longer simulation times, especially for complex or large-scale simulation scenarios.

c) *Channel representation*

Accurate modeling of wireless communication channels can be challenging in discrete event-based simulators. The accuracy of channel representation can impact simulated performance and often requires experimental validation to ensure result fidelity.

B. Queueing Theory-based Simulators

Queueing theory-based simulators are another widely used category of simulators for evaluating the performance of wireless networks. These simulators rely on queueing theory principles to model the behavior of network nodes and analyze performance in terms of throughput, latency, and quality of service.

1) Characteristics of Queueing Theory-based Simulators

a) *Queueing mode*

These simulators employ queueing models to represent network nodes. Nodes are treated as servers that receive requests (packets) and process them according to specific policies. Queues are used to manage packet arrivals and departures Fig. 2.

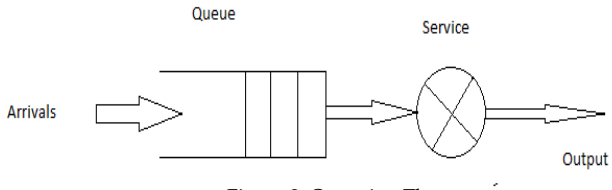


Figure 2. Queueing Theory

b) *Queueing theory*

Queueing theory-based simulators leverage the principles of this mathematical discipline to analyze and predict network performance. They use mathematical models to calculate performance metrics such as average waiting time, throughput, probability of loss, etc.

c) *Variety of policies*

These simulators allow defining different queueing policies, such as First-In-First-Out (FIFO) queue, Priority queue, Round Robin queue, etc. This enables evaluating the impact of queue management policies on network performance.

2) Advantages

a) *Mathematical analysis*

Queueing theory allows for in-depth mathematical analysis of network performance. The obtained results can be used to predict performance in real-world scenarios and make informed decisions regarding network improvements.

b) *Flexibility*

These simulators offer great flexibility to model different types of queues, processing policies, and network scenarios. They can be used to evaluate wireless network performance under various conditions, such as varying traffic loads, different throughputs, etc.

c) *Reduced simulation time*

Queueing theory-based simulators are often faster than event-driven simulators. By employing analytical models, they allow for obtaining results more quickly, which can be useful when evaluating multiple scenarios or searching for optimal solutions.

3) Disadvantages

a) *Model simplifications*

Queueing theory-based simulators rely on simplified models to represent network behavior. These models may not capture all the details and complex interactions of the real network, which can lead to less accurate results.

b) *Assumption dependency*

These simulators depend on assumptions made about queueing behavior, such as packet arrival following a specific distribution, queue stability, etc. If these assumptions do not match real-world conditions, the results can be biased.

4) Examples of queueing theory-based simulators

a) *QualNet*

QualNet is a commercial simulator that includes queueing theory-based models to evaluate wireless network performance *[4]*. It offers a wide range of features to model queues, processing policies, routing protocols, etc., enabling detailed performance analysis.

C. Emulation Simulators

Emulation simulators are another category of simulators used in evaluating the performance of wireless networks. Unlike simulators based on mathematical models or discrete events, emulation simulators more faithfully reproduce the real behavior of networks by using real hardware and software equipment [5]. Emulation simulators utilize real hardware such as routers, switches, Wi-Fi access points, etc., to replicate real conditions of a wireless network. This allows for more realistic and accurate results by simulating real-world behaviors.

1) Characteristics of Emulation Simulators

a) *Real-world experiments*

Emulation simulators enable conducting real-world experiments using real applications, protocols, and equipment. This allows for evaluating performance under real conditions and reproducing specific scenarios that may not be easily modeled or simulated by other types of simulators.

b) *Flexibility and control*

These simulators offer great flexibility and precise control over experiments. Users can configure network parameters, equipment behaviors, traffic loads, etc., to replicate specific scenarios and observe real-time performance.

2) Advantages of emulation simulators

a) *Realism*

Emulation simulators provide a high level of realism by replicating real conditions of wireless networks. This allows for more precise performance

evaluation, taking into account real interactions between equipment and protocols.

b) *Solution validation*

These simulators allow for validating real solutions using existing equipment and protocols. This enables testing protocol updates, deployment strategies, network configurations, etc., before deploying them in real-world environments. Problem detection: Emulation simulators enable the detection and diagnosis of potential issues in wireless networks. By replicating real conditions, they allow for observing and analyzing performance under different traffic loads, topologies, etc., to identify bottlenecks and optimize performance.

3) Disadvantages of emulation simulators

a) *Cost*

Emulation simulators can be expensive due to the use of real hardware. Acquiring and maintaining equipment can represent a significant investment.

b) *Complexity*

Due to the real nature of the equipment and configurations, emulation simulators can be more complex to set up and use compared to other types of simulators. Technical expertise is often required to effectively operate them.

4) Examples of emulation simulators

a) *GNS3 (Graphical Network Simulator)*

Is a popular emulation simulator used for modeling and testing wireless networks *[6]*. It allows for connecting real equipment or virtual images to create complex networks and conduct real-world experiments.

b) *CORE (Common Open Research Emulator)*

Is another widely used emulation simulator in network research. It enables creating complex network topologies and simulating wireless networks using virtual and real equipment *[7]*.

These simulators represent just a few examples of the available tools. Each has its own advantages and limitations, and the choice depends on specific research or application needs. In the next section, we will provide detailed information about our simulator model used to evaluate the performance of 802.11be in high-density scenarios.

III. OUR SIMULATION APPROACH

The simulator developed aims to simulate the behavior of a real wireless network, which consists of a collection of connected devices. The main objective is to simulate the transmission, propagation, and reception of electromagnetic signals. The simulator is designed to faithfully replicate the interactions between devices in a wireless network. It takes into account various stages of communication, including signal generation, wave propagation in the surrounding space, potential interferences with other signals, and the reception of the signal by receiving devices.

A. *Principle of Operation of Our Simulator*

Our wireless network simulator is built on a straightforward and adaptable architecture, designed to facilitate the exploration and visualization of various wireless network scenarios. The fundamental approach of our simulator is based on the creation of sophisticated simulation models that faithfully replicate the behaviors of wireless networks. What sets it apart is its ability to enable researchers to test and apply their ideas from the early stages of research. Unlike many other simulators, which are typically employed towards the end of the research process for solution validation, ours encourages an exploratory approach. It allows researchers to experiment with different values during simulations to observe how the network responds and to identify the advantages and disadvantages of their concepts, thereby fostering innovation and continuous improvement of their ideas.

B. *Differences Compared to Other Simulators*

The primary distinction between our simulator and other existing solutions lies in its early use at the onset of research work. In contrast to the majority of simulators, traditionally reserved for validation purposes in advanced stages, ours promotes exploration and discovery by allowing researchers to test their ideas right from the beginning. This unique approach nurtures creativity by providing a flexible platform for experimentation. It offers an innovative solution for research and development in the field of wireless networks, emphasizing the generation of new ideas and the discovery of innovative solutions.

Our simulator is built upon a modular architecture, consisting of two essential parts, each playing a fundamental role in the realistic representation of wireless networks.

C. Physical Part: Simulation of Radio Waves

The first part, the physical component, is dedicated to accurately simulating the behavior of radio waves. It includes several key components, including:

1) *Wave Class*

This class is responsible for generating and detecting radio waves. It encapsulates crucial properties such as frequency (f), power (P), phase (ϕ), amplitude (A), and a timer (t) for recording the arrival of waves at specific locations.

2) *Antenna Class*

The Antenna class generates and detects radio waves using the Wave object. It has two essential attributes, ChannelOut (where generated waves are stored) and ChannelIn (a list of incoming or future radio waves).

3) *Modulation and Demodulation*

The Modem class ensures the modulation of data into radio signals and their inverse demodulation. It facilitates the conversion of bits into signals and vice versa using defined coding schemes.

4) *Propagation of Radio Wave*

This feature simulates the propagation of radio waves through the wireless channel, taking into account realistic reflections (R) and diffractions (D) *[8]*. Each radio wave undergoes random alterations in amplitude, phase, and power, based on its distance (d) from the transmitter. The speed of light in a vacuum (c = 300,000 km/s) is used to calculate time delays.

5) *Interference*

Interference between radio waves is accurately modeled. The following equations fig. 1 and fig. 2 describe the characteristics of the resulting wave during interference between two initial waves:

Resulting Phase:

$$\text{Result\_}\varphi = \varphi 1 + \varphi 2 + 2\pi f(t1 - t2) \quad (1)$$

The resultant phase represents the final phase of a wave resulting from the interference of two or more initial waves. When multiple waves overlap or interfere with each other, their individual phases combine to produce a new phase for the resultant wave.

$\text{Result\_}\varphi$: The resultant phase.

ϕ_1 and ϕ_2: The initial phases of the interfering waves.

f: The frequency of the waves.

t1 and t2: The respective times at which the interfering waves reach a specific point in the simulation.

The phase represents the position of the wave at a specific point in time. When two or more waves overlap, their individual phases are summed, along with the phase shift due to the time difference ($2\pi f(t1 - t2)$), to determine the phase of the resultant wave.

Resulting Amplitude:

$$\text{Result\_}A = \sqrt{(A1\text{^}2 + A2\text{^}2 + 2A1A2\cos(\varphi 1 - \varphi 2))} \quad (2)$$

Result_A: The resultant amplitude.

A1 and A2: The initial amplitudes of the interfering waves.

The resultant amplitude (Result_A) represents the final amplitude of a wave resulting from the interference of two or more initial waves. When waves interfere, their individual amplitudes combine to produce a new amplitude for the resultant wave.

The resultant amplitude is determined by considering both the individual amplitudes and the phase difference between the interfering waves. The term $2A_1A_2\cos(\phi_1 - \phi_2)$ accounts for the constructive or destructive interference between the waves. When the phase difference results in constructive interference, the amplitudes add up; when it results in destructive interference, the amplitudes partially or completely cancel each other out.

These formulas are fundamental in understanding how waves interact and combine during interference, a crucial aspect of simulating realistic wireless communication scenarios in your simulator.

D. *Computer Part: Data Management*

The second part, the computer component, focuses on data management, their conversion into bits, and the detection and correction of transmission errors. This part is dominated by the DataLink class, which combines the MAC and LLC layers. It integrates crucial elements:

1) *MAC Address*

Each instance of the DataLink class is associated with a unique MAC address, allowing the identification of devices in the network.

2) *Random Data/Destination Generator*

Each device is equipped with a random data and destination generator. This feature simulates realistic data exchanges between network devices.

The DataLink class implements essential methods to ensure effective data link management.

3) *framing()*

This method forms frames from the pending bit list, following defined protocols (such as the 802.11 protocol).

4) *sensing()*

The sensing() function monitors the channel before transmission to detect the presence of other signals.

5) *send()*

Is responsible for sending modulated data using the modulate() method of the Modem class.

6) *receive()*

It processes the data received by the Modem, performing inverse demodulation to obtain the bits.

7) *inspect()*

This function inspects received frames to detect potential transmission errors and apply corrections if necessary.

E. *Parallelism Management*

One of the major challenges in simulating wireless networks in a single-processor environment is managing parallelism. To realistically replicate independent communication operations between devices, our simulator uses a meticulously designed mechanism called a "parallel simulation cycle."

1) Parallel Simulation Cycle

The simulation cycle represents a unit of time in our simulator, equivalent to one nanosecond of simulation time. During each cycle, all instructions in the simulation program are executed once, but in an ordered and synchronized manner. This ensures that tasks are executed in a logical order, mimicking the parallelism of communication operations in a real wireless network. To illustrate how parallel simulation cycle works, consider a logical sequence of tasks for wireless transmission Fig. 3:

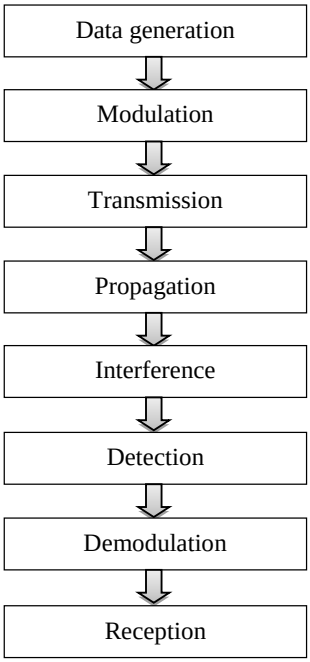


Figure 3. Logical sequence of wireless transmission tasks

In a real environment, these tasks are executed in parallel and are completely independent of each other. To replicate this behavior, the simulation cycle is structured as follows:

1-> Execute the reception function in each device.

2-> Execute the demodulation function in each device.

3-> Execute the detection function in each device.

4-> Execute all interference operations.

5-> Execute all propagation operations.

6-> Execute the transmission function in each device.

7-> Execute the modulation function in each device.

8-> Execute the data generation function in each device

At the end of this simulation cycle, one nanosecond of simulation time has passed. This process ensures that tasks are executed in a realistic order while maintaining precise synchronization between operations Fig. 4. Thus, data generated in one cycle will be modulated in the next cycle and transmitted in the cycle after that, faithfully reflecting the reality of

wireless networks where many parallel operations occur simultaneously.

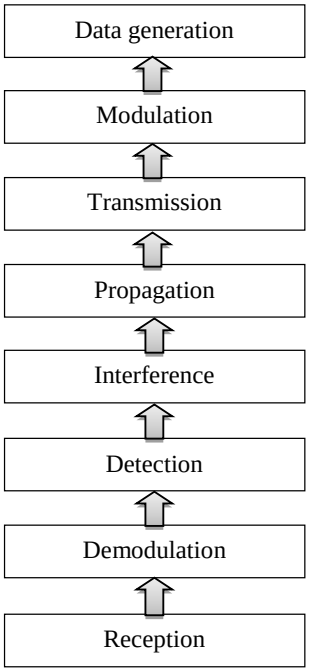


Figure 4. Logical execution order for the parallel tasks

F. Main Simulator Classes

In addition to the previous simulation components, our simulator comprises two essential main classes: the "Device" class and the "Simulator" class, responsible for managing individual devices and the overall simulation.

1) Device Class

The "Device" class represents a complete network card and contains several essential components for modeling a device in the wireless network. It includes the following elements:

a) *Device Location*

A device can be located at a specific location, and it can also be in motion. Location is crucial for simulating device mobility within the network.

b) *Antennas*

One or more antennas are associated with each device. Antennas are responsible for generating and detecting radio waves, contributing to communication within the network.

c) *Modem*

Each device is equipped with a modem, which plays a central role in modulation and demodulation of data, transforming them into radio signals and vice versa.

d) *Data Link Layer*

The data link layer is managed by the "DataLink" class described earlier. It ensures data management, from transmission to reception, while also guaranteeing error detection and correction.

e) *Random Data/Destination Generator*

Each device is equipped with a random data and destination generator. This component is crucial for simulating realistic data exchanges between network devices.

Additionally, the "Device" class has the following methods:

a) *Send()*

This method is used to send data across the network, utilizing the appropriate components such as the antenna, modem, and data link layer.

b) *Receive()*

It is responsible for receiving bits at the destination, managing signal demodulation and handling received data.

c) *Move()*

This method allows for device movement, taking into account its current location and velocity if the device is in motion.

2) Simulator Class

The "Simulator" class is the main class of our simulator, responsible for managing the overall simulation of the wireless network. It contains the following elements:

a) *Number of Devices*

This is the total number of devices in the network that we wish to simulate.

b) *List of Devices*

This list is initialized with the required number of devices to form the network. Each device is created based on defined parameters, such as location, antennas, modem, data link layer, and random data/destination generator.

c) *Clock*

The simulation clock is initialized at 0, with a time unit in nanoseconds. It is essential for synchronizing actions and events in the simulation.

d) *Max Simulation Time*

This parameter determines the total duration of the simulation, specifying when the simulation should end.

The "Simulator" class coordinates the execution of the simulation, managing device movement, data transmission and reception, and the progression of time until the maximum simulation duration is reached.

These two main classes, "Device" and "Simulator," form the foundation of our wireless network simulator, allowing for realistic modeling of individual devices and network dynamics as a whole.

In summary, while state-of-the-art simulators have proven themselves as robust tools for network simulation, our simulator distinguishes itself by offering a specialized focus on wireless networks, a modular architecture, realistic wireless wave propagation, sophisticated parallelism management, and a strong emphasis on customization for research purposes. Researchers and developers looking for a simulation tool tailored to wireless scenarios with unique features will find our simulator a valuable addition to their toolkit.

IV. SIMULATION OF 802.11BE

We conducted an extensive evaluation of the 802.11be standard using our simulator, focusing on its performance in high-density scenarios. Our simulations involved varying numbers of devices, ranging from 2 to 128, and aimed to assess the impact of density on the network's throughput and latency.

A. *Throughput Analysis*

In terms of throughput, the results of our simulations Fig. 5 were impressive. With just 2 devices, we observed a staggering throughput of 30.6 Gbps, showcasing the exceptional capacity of 802.11be. As the density increased, the network continued to deliver substantial throughput. In scenarios with 8, 16, and 32 devices, the throughput values were 3.8 Gbps, 1.92 Gbps, and 950 Mbps, respectively. Even in densely populated networks with 64 and 128 devices, the throughput remained significant at 400 Mbps and 200 Mbps, respectively. These results highlight the ability of 802.11be to maintain high data rates in challenging high-density environments.

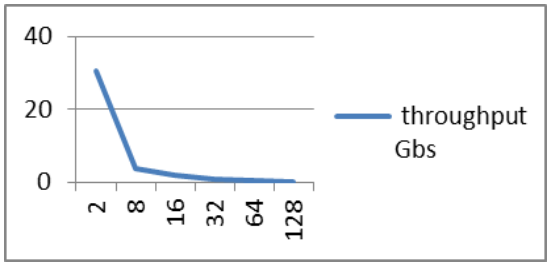


Figure 5. Impact of Density on Throughput Performance in 802.11be

B. *Latency Analysis*

When evaluating latency, 802.11be demonstrated remarkable performance across all density scenarios Fig. 6. With just 2 devices, the average latency was measured at 51 microseconds, indicating swift data transmission and minimal delays. As density increased, average latency remained consistently low: 450 microseconds (8 devices), 830 microseconds (16 devices), and 1.6 milliseconds (32 devices). Even in heavily congested networks with 64 and 128 devices, latency values were acceptable at 3.3 milliseconds and 6.5 milliseconds, respectively.

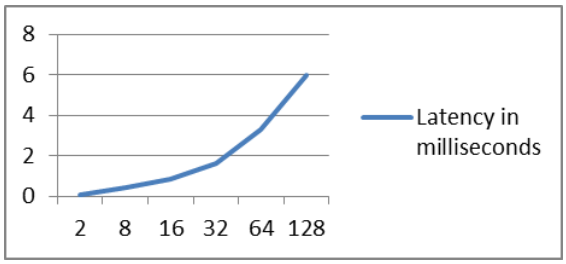


Figure 6. Impact of Network Density on Latency in 802.11be

C. *Robustness in High-Density Environments*

The simulations revealed the exceptional robustness of 802.11be in high-density scenarios. Despite the increasing number of devices, the network maintained reliable and efficient performance. This can be attributed to advanced techniques like OFDMA (Orthogonal Frequency-Division Multiple Access), enabling efficient resource allocation and minimizing interference in crowded environments. Additionally, MU-MIMO (Multi-User Multiple Input, Multiple Output) enhances spectral efficiency and overall network capacity, allowing multiple devices to communicate simultaneously without compromising performance

In conclusion, our simulation results confirm that 802.11be is a highly capable standard for high-density deployments. Its impressive throughput, low latency, and robust performance in crowded environments position it as a promising solution for future wireless networks. By delivering high data rates and minimizing delays, 802.11be ensures an enhanced user experience even in scenarios with a large number of connected devices. The successful evaluation of 802.11be in our simulations opens up exciting possibilities for its practical implementation and future enhancements. Some potential perspectives to explore include:

1) Real-world Deployment

Validating the simulation results through real-world deployment and testing would provide valuable insights into the actual performance of 802.11be in high-density environments.

2) Energy Efficiency

Optimizing the energy efficiency of 802.11be through power-saving mechanisms and dynamic power control strategies is important for sustainability.

In summary, the positive simulation results highlight the potential of 802.11be in high-density scenarios. Ongoing efforts in performance optimization, interference mitigation, energy efficiency, security enhancements, and standard evolution will further establish 802.11be as a key wireless networking technology.

REFERENCES

[1] G. F. Riley and T. R. Henderson, "The ns-3 Network Simulator," in Modeling and Tools for Network Simulation," Springer, 2010, pp. 15-34. https://doi.org/10.1007/978-3-642-12331-3_2.

[2] A. Varga, "OMNeT++," in Modeling and Tools for Network Simulation," K. Wehrle, M. Güneş, and J. Gross, Eds. Berlin, Heidelberg: Springer, 2010, pp. 35-59. https://doi.org/10.1007/978-3-642-12331-3_3.

[3] Z. Lu and H. Yang, "Unlocking the Power of OPNET Modeler. Cambridge University Press," 2012, 254 pages.

[4] K. P. AL-Sakib, M. M. Muhammed, and K. Shafiullah, "Simulation Technologies in Networking and Communications: Selecting the Best Tool for the Test," Boca Raton: CRC Press 2014.

[5] John B. Copp "The COST Simulation Benchmark: Description and Simulator Manual" Directorate-General for Research, Belgique, 2002.

[6] L. N. Dayanand, B. Ghorbani, and S. Vaghri, "A Survey on the Use of GNS3 for Virtualizing Computer Networks," International Journal of Computer Science and Engineering, 2016, 5(1), pp 49-58.

[7] S. Tan, W. -Z. Song, Q. Dong, and L. Tong, "SCORE: Smart-Grid Common Open Research Emulator,"IEEE Third International Conference on Smart Grid Communications (SmartGridComm), Tainan, Taiwan, 2012, pp. 282-287.

[8] G. de La Roche, "Simulation de la Propagation des Ondes Radio en Environnement Multi Trajets pour l'Étude des Réseaux Sans Fil" PhD thesis, INSA de Lyon, 2007.

[9] K. Wehrle, M. Günes and J. Gross. "Modeling and Tools for Network Simulation" Media, 2010.

A Finite Markovian Queue with Impatient Customers Under Triadic Policy: Reliability Measures

Kadi Abir
Laboratory of Applied Mathematics, Department of Mathematics, Bejaia University, 06000 Bejaia, Algeria.
abir.kadi@univ-bejaia.dz

Touche Nassim
Research Unit LaMOS University of Bejaia, 06000 Bejaia, Algeria.
nassim.touche@univ-bejaia.dz

Boualem Mohamed
Research Unit LaMOS University of Bejaia, 06000 Bejaia, Algeria.
mohammed.boualem@univ-bejaia.dz

***Abstract*—**
The main purpose of this paper is to investigate the performance analysis of $M/M/2$ machine repair problem with finite capacity L operating under the triadic policy $(0, Q, N, M)$, impatience and working vacation policy. As soon as the system becomes empty, both servers leave for working vacations wherein only one of the two servers provide a service during the vacation. Steady state probabilities that describe the number of failed machines in system are derived and taken in closed form. Different performance measures of the system are developed and analyzed with numerical illustrations to investigate the reliability of the model.

***Keywords*—**
Queueing system, vacation, Markov chain, impatience, machine repair problem.

I. INTRODUCTION

The interaction between humans and machines in industrial plants and other dynamic technical systems is crucial for ensuring performance quality and efficiency. Whenever machines are present, they are susceptible to failure, necessitating repairs. To achieve optimal system performance, it is imperative to conduct a scientific study on the interaction between service providers (servers) and customers. Servers may become temporarily unavailable for various reasons, a concept known as vacation in queueing theory.

In such systems, it is common to observe customers arriving at a queue but deciding not to join, a behavior referred to as balking. Customers may choose not to join the queue for various reasons. Conversely, even when a customer decides to join the queue, they may become impatient and leave before receiving service, which is known as "reneging." In the context of a machining system, both balking and reneging can be observed in the actions of the caretaker responsible for repairing failed machines.

Queueing system has an important role in predicting queues features in many domain of daily life, as well as manufacturing systems, production systems, call centers and many various domains. We are interested in this work of machine repair system. Machine repair system are widely used in common industrial fields and techno-socio advances.

Vacation queues and customers' impatience have attracted many researchers because of their wide applications in many real life such as inventory management, production optimization, manufacturing systems, transportation. Variant vacation queueing system with Bernoulli feedback, balking and server's states-dependent reneging was showed by Bouchentouf et al. [1] using generating functions to obtain the steady state solution. Also, Bouchentouf et al. [2] analyzed a multi-server model with finite capacity, multiple synchronous working vacations and balking by the matrix-geometric methode. Machining systems are very important in many domain of real-life. The failure of machining system is quite common on service sectors, computer and telecommunication networks. Shekhar et al. [5] studied a model in which the number of operating machines and the warm standby machines are finite under the control of a single unreliable server. N and vacation policies are also considered. Machine repair model with standbys, working vacation and server breakdown was showed by Jain and Preeti [7]. In Bhagat et al. [3] paper's a retrial machine repair system with M identical operating machines was studied under single repairman, controled arrival, the F-policy working vacation, breakdown and repair problem. Rhee [6] was the first author who introduced triadic policy for the $M/M/2$ queueing system model.

Ketema et al. [9] applied the triadic policy $(0, Q, N, M)$ for the $M/M/2$ machine repair problem and determined the optimal working vacation service rate η^* and the optimal operating triadic parameters $(0, Q^*, N^*, M^*)$. Lin et al. [4] considered an M/M/2 with infinity capacity. Analytic closed-form solutions of the

queueing system operating under the triadic $(0, Q, N, M)$ policy are derived. And then developed the total expected cost function per unit time, to obtain the optimal operating $(0, Q, N, M)$. The impatient customers for the machine repair system was studied for the first time by Ketema. [8] considered an $M/M/2$ machine repair system with multiple working vacations (MWV), impatient customers and triadic policy $(0, Q, N, M)$.

The rest of the paper is organized as follows. In Section 2, we introduce the mathematical description of the proposed queueing model. In Section 3, the set of the balance equations and transitions matrix are presented the steady-state results of the queueing model is provided. In Section 4, we give a real example for the proposal model. Performance measurement represented in section 5. A numerical simulation results showed at the end of the work.

II. MODEL DESCRIPTION

We consider a machine repair model with capacity L operating machines maintained by a two repairman. The assumptions of the model are built up as follows:

1) The capacity of the system is assumed to be a finite number L operating machine.
2) If the operating machine fails it join the system to repair inter-arrivals for the failed machine accrue according to a exponential process with rate λ.
3) Failed machine decides either to join the queue with probability β_i, or balk with probability $1-\beta_i$, for $0 \leq i \leq L$ and where: $\beta_0 = 1$, $0 < \beta_{i+1} \leq \beta_i \leq 1, 1 \leq i \leq L-1$ and $\beta_L = 0$.
4) Service follows an exponential process with rate μ in busy period and μ_v in the vacation period ($\mu_v < \mu$).
5) When the failed machine enter the system, it activates a timer T follows an Exponential process with rate: ξ in the dormant working vacation period, failed machine leaves the queue with probability α, he can return to the system with probability $1-\alpha$.
6) The triadic policy:
   - It takes the systemic bellow:
     - ◇ When the number of waiting machines reaches N, only one server will start the busy period instantly.
     - ◇ After a while, if the number of waiting machines increases to level M $(M > N)$, then the second server becomes active.
     - ◇ If the number of machines in the system decreases to Q $(Q < N)$, while both servers are active simultaneously, the server just finishing service becomes inactive.

   This policy is known as a triadic $(0, Q, N, M)$ policy. In addition, if the number of machines in the system increases to zero when one server is active, then all servers start a working vacation period (WV).
7) Vacation durations assumed to be exponentially distributed with rates ϕ.

We assume that inter arrival times, service times, and vacation times are mutually independent. In addition, the service order is First-In-First-Out (FIFO). The following notations are used in further analysis of the model, for $0 \leq n \leq L-1$:

L : The number of operating machines,
$(L-n)\lambda$: Arrival rate of failed machines.

III. ANALYSIS OF THE STEADY-STATE PROBABILITY DISTRIBUTION

We can define the quasi birth and death process (QBD) of $\{N(t), J(t)\}$ with the state space:

$$\Omega = \{(0,0) \cup (n,j) : 0 \leq n \leq L, j = 0,1,2,3\}.$$

Let $N(t)$ be the number of customers in the system at time t and let:

$$J(t) = \begin{cases} 0, & if\ one\ server\ is\ in\ \text{WV}\ period, \\ 1, & if\ only\ one\ server\ is\ turned\ on\ and\ active, \\ 2, & if\ both\ servers\ are\ turned\ on, \\ 3, & if\ both\ servers\ are\ turned\ off. \end{cases}$$

We denote that for $0 \leq i \leq L$:

$$\lambda_i = (L-i)\lambda\beta i,$$
$$\zeta_i = i\alpha\xi.$$

A. Steady state equations

The balance equations of the model are:
For $J(t) = 0$:

$$(\lambda_0 + \phi)P_{0,0} = \mu_v P_{0,1} + \mu P_{1,1}, n = 0, \quad (1)$$

$$[\lambda_n + \mu_v + \zeta_{n-1} + \phi]P_{0,n} = \lambda_{n-1}P_{0,n-1} + (\mu_v + \zeta_n)P_{0,n+1}, \quad 1 \leq n \leq N-1, \quad (2)$$

$$[\lambda_n + \mu_v + \zeta_{n-1} + \phi]P_{0,n} = \lambda_{n-1}P_{0,n-1} + (\mu_v + \zeta_n)P_{0,n+1}, \quad N \leq n \leq L-1, \quad (3)$$

$$[\zeta_{L-1} + \mu_v + \phi]P_{0,L} = \lambda_{L-1}P_{L-1}, \quad n = L. \quad (4)$$

For $J(t) = 1$:

$$[\mu + \lambda_1]P_{1,1} = (\mu + \zeta_1)P_{1,2}, \ n = 1, \quad (5)$$

$$[\lambda_n + \mu + \zeta_{n-1}]P_{1,n} = \lambda_{n-1}P_{1,n-1} + (\mu + \zeta_n)P_{1,n+1}, \quad 2 \leq n \leq Q-1, \quad (6)$$

$$[\lambda_Q + \mu + \zeta_{Q-1}]P_{1,Q} = \lambda_{Q-1}P_{1,Q-1} + (\mu + \zeta_Q)P_{1,Q+1} + \quad Q \leq n \leq N-1, \quad (7)$$

$$[\lambda_n + \mu + \zeta_{n-1}]P_{1,n} = \lambda_{n-1}P_{1,n-1} + (\mu + \zeta_n)P_{1,n+1}, \quad N \leq n \leq M-2, \quad (8)$$

$$[\lambda_{M-1} + \mu + \zeta_{M-2}]P_{1,M-1} = \lambda_{M-2}P_{1,M-2}, \ n = M-1. \quad (9)$$

For $J(t) = 2$:

$$[\lambda_{Q+1} + 2\mu]P_{2,Q+1} = [2\mu + \zeta_{Q+1}]P_{2,Q+2}, \ n = Q+1, \tag{10}$$

$$[\lambda_n + 2\mu + (n-1)\alpha\xi_1]P_{2,n} = \lambda_{n-1}P_{2,n-1} + (2\mu + \zeta_n)P_{2,n+1}, \ Q+2 \le n \le N-2, \tag{11}$$

$$[\lambda_n + 2\mu + \zeta_{n-1}]P_{2,n} = \lambda_{n-1}P_{2,n-1} + (2\mu + \zeta_n)P_{2,n+1} + \phi P_{0,n}, \ N-1 \le n \le M, \tag{12}$$

$$[\lambda_n + 2\mu + \zeta_{n-1}]P_{2,n} = \lambda_{n-1}P_{2,n-1} + (2\mu + \zeta_n)P_{2,n+1} + \phi P_{0,n}, \ M+1 \le n \le L-1, \tag{13}$$

$$[2\mu + \zeta_{L-1}]P_{2,L} = \lambda_{L-1}P_{2,L-1} + \phi P_{0,L}, \ n = L. \tag{14}$$

For $J(t) = 3$:

$$L\lambda_0 P_{3,0} = \phi P_{0,0}, \ n = 0, \tag{15}$$

$$[\lambda_n + \zeta_{n-1}]P_{3,n} = \lambda_{n-1}P_{3,n-1} + \zeta_n P_{3,n+1} + \phi P_{0,n}, \ 1 \le n \le N-2, \tag{16}$$

$$[\lambda_{N-1} + \zeta_{N-2}]P_{3,N-1} = \lambda_{N-2}P_{3,N-2} + \phi P_{0,N-1}, \ n = N-1. \tag{17}$$

IV. PRACTICAL EXAMPLE

The queueing system studied the triadic policy could appear in the injection therapy room in a hospital.

There are two nurses (servers) operates the chemotherapy clinic, which supports patients from breast thyroid surgery.There are L injection seats. The L injection seats can be considered as L operating machines. As a patient arrives at the chemotherapy clinic, the nurse leads the patient to take an injection seat for the chemotherapy it considered as a working vacation with rate: μ_v. This situation, a seat with a patient waiting for service, can be considered as a failed machine requiring the server to repair it. The nurse will wait for the other arriving patients until the number of patients is over N and then start the chemotherapy (start the busy period with one server). At a later time, when the number of patients waiting for service increases to the specific quantity of M, the other nurse will leave the reception desk temporarily and join to the injection service to reduce the waiting time for patients. However, when the number of patients in the clinic decreases to Q $(Q < N)$, one nurse will remove from the patient service and return to the reception desk to prepare documents and notify the family member of the patient for the following process this act also considered us a working vacation epoch for the second nurse.

V. PERFORMANCE MEASURES

A. Expected value

1) The expected number of failed machines in the system:

$$E[N] = \sum_{n=0}^{L} n\pi_{0,n} + \sum_{n=1}^{M-1} n\pi_{1,n} + \sum_{n=Q+1}^{L} n\pi_{2,n} + \sum_{n=1}^{N-1} n\pi_{3,n}.$$

2) The expected number of operating machines in the system:

$$E[OP] = L - E(N).$$

3) The expected number to have only one server is busy during the WV period:

$$E[B_0] = \sum_{i=0}^{L} \pi_{0,n}. \tag{18}$$

4) The expected number to have only one server is busy during the regular busy period.

$$E[B_1] = \sum_{i=1}^{M-1} \pi_{1,n}. \tag{19}$$

5) The expected number to have both servers are busy during the regular busy period.

$$E[B_2] = 2\sum_{i=Q+1}^{L} \pi_{2,n}. \tag{20}$$

6) The expected number to have only one server is busy server during the dormant period:

$$E[B_3] = \sum_{i=0}^{N-1} \pi_{3,n}. \tag{21}$$

7) The expected queue length:

$$E_Q = \sum_{i=1}^{L}(n-1)\pi_{0,n} + \sum_{i=1}^{M-1}(n-1)\pi_{1,n} + \sum_{i=Q+1}^{L}(n-2)\pi_{2,n} + \sum_{i=1}^{N-1}(n-1)\pi_{3,n}. \tag{22}$$

8) The expected number of idle servers in the system:

$$E_i = 2\pi_{0,0} + \sum_{i=1}^{L}\pi_{0,n} + \sum_{i=1}^{M-1}\pi_{1,n} + 2\sum_{i=1}^{N-1}\pi_{3,n}. \tag{23}$$

B. Reliability measures

1) Machine availability:

$$AV = \frac{E[OP]}{L}. \tag{24}$$

2) Operative utilization (the fraction of busy servers):

$$BS = \frac{E[B_0] + E[B_1] + E[B_2] + E[B_3]}{2}. \tag{25}$$

3) The average balking rate:

$$BR = \sum_{i=1}^{L}(n-1)\lambda(1-b_n)\pi_{0,n} + \sum_{i=1}^{M-1}(n-1)\lambda(1-b_n)\pi_{2,n} + \sum_{i=Q+1}^{L}(n-2)\lambda(1-b_n)\pi_{2,n} + \sum_{i=1}^{N-1}(n-1)\lambda(1-b_n)\pi_{3,n}. \tag{26}$$

4) The average reneging rate:

$$AR = \xi \sum_{i=1}^{L}(n-1)\pi_{0,n} + \xi \sum_{i=1}^{M-1}(n-1)\pi_{1,n} + \xi(\sum_{i=Q+1}^{L}(n-2)\pi_{2,n} + \xi \sum_{i=1}^{N-1}(n-1)\pi_{3,n}). \quad (27)$$

5) The average costumer loss rate:

$$LR = BR + AR. \quad (28)$$

VI. NUMERICAL RESULTS

The steady state probabilities of all system states of the machine repair problem are computed in this section. The matrix-geometric in it's recursive approach is employed to evaluate these probabilities. The numerical results are facilitated to explore the effects of various parameters on the system performance. For the computation purpose, the program for geometric-matrix method is coded in R-software. In order to compute the performance indices, we set default parameters as: $L = 32, M = 25, N = 19$, $Q = 8$, $\lambda = 1.9$, $\mu = 2$, $\mu_v = 1.5$, $\xi = 2$, $\alpha = 0.4$, $b = 0.9$ and $\phi = 2$.

Figure 1 represents the effect of β on (AR) an (BR):

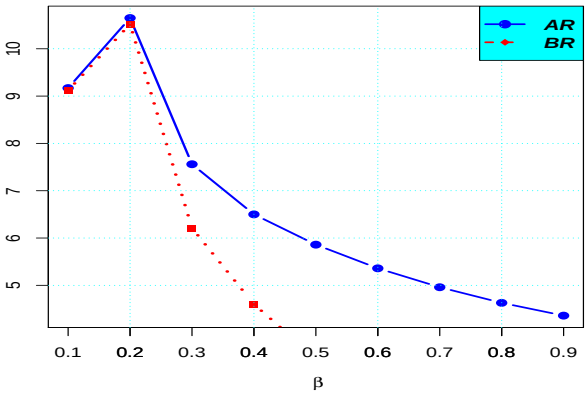


Figure 1. The effect of β on AR and BR.

For λ takes three different values: $(0.7, 1.5, 2)$, and $\alpha \in [0.1-0.9]$. Figure 2 represents the effect of and α on (LR):

For different values of Q, N and M. Figure 3 represents the effect of λ on E_q:

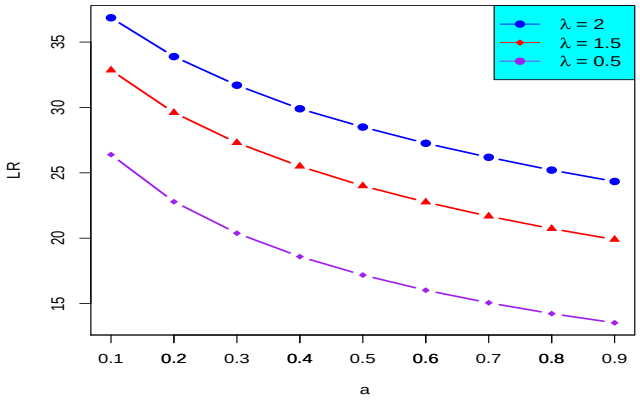


Figure 2. The effect of λ and α on LR.

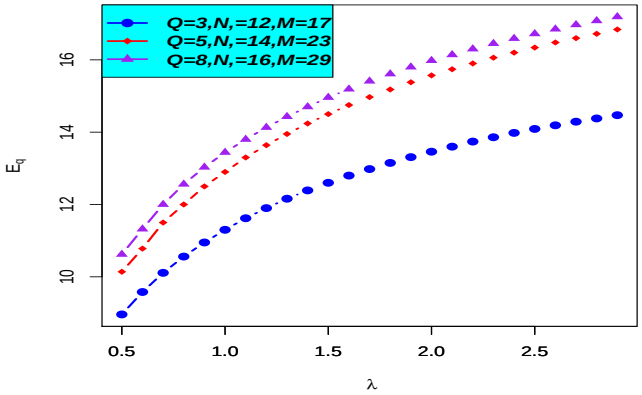


Figure 3. The effect of λ on E_q.

VII. CONCLUSION

In this paper, we considered an $M/M/2$ queue under working vacation and customers impatient during working vacation period and when the service is off. The steady-state equations were obtained. The reliability measures and some numerical investigations for the model were showed at the end. Solution of the steady-state solution of the equations will be showed in the topic of the further work.

References

[1] A. A. Bouchentouf, M. Boualem, M. Cherfaoui, L. Medjahri, "Variant vacation queueing system with Bernoulli feedback, balking and server's states-dependent reneging," *Yugoslav Journal of Operations Research 31.* no. 4 (2021): 557–575.

[2] A. A. Bouchentouf, M. Boualem, L. Yahiaoui, H. Ahmad, "A multi-station unreliable machine model with working vacation policy and customers' impatience," *Quality Technology & Quantitative Management 19.* no. 6 (2022): 766–796.

[3] A. Bhagat, S. Rachita, G. Deepika, "Controlled Arrival Machine Repair Problem with Working Vacation and Reattempts", *International Journal of Mathematical. Engineering and Management Sciences.*6.1 (2021): 279.

[4] C.H. Lin, J.C. Ke, "Genetic algorithm for optimal thresholds of an infinite capacity multi-server system with triadic policy", *Expert Systems with Applications 37.* no. 6 (2010): 4276–4282.

[5] C. Shekhar, P. Deora, S. Varshney, K. P. Singh, D. C. Sharma, "Optimal profit analysis of machine repair problem with repair in phases and organizational delay," *International journal of mathematical, engineering and management sciences 6.* no. 1 (2021): 442.

[6] H. K. Rhee, B. D. Sivazlian, "Distribution of the busy period in a controllable $M/M/2$ queue operating under the triadic $(0, K, N, M)$ policy", *Journal of Applied Probability 27.* no. 2 (1990): 425–432.

[7] M. Jain, Preeti, "Cost analysis of a machine repair problem with standby, working vacation and server breakdown", *International Journal of Mathematics in Operational Research 6.* no. 4 (2014): 437–451.

[8] T. Ketema, "Performance Analysis of Machine Repair System with Balking, Reneging, Multiple Working Vacations and Two Removable Servers Operating under the Triadic $(0, Q, N, M)$ Policy", *Innovative Systems Design and Engineering 11.* no.4 (2020).

[9] T. Ketema, D. Seleshi, M. T. Belachew, "Controllable $M/M/2$ machine repair problem with multiple working vacations and triadic (0, Q, N, M) policy", *International Journal of Management Science and Engineering Management 16.* no. 3 (2021): 184–196.

Performance Study of Up-Link OFDMA Random Access for IoT Applications-based WiFi 7

Mammeri Souhila
LaMOS Laboratory
Faculty of Technology
University of Bejaia
souhila.mammeri@univ-bejaia.dz

Ould Amara Said
LaMOS Laboratory
Faculty of Exact Sciences
University of Bejaia
said.ouldamara@univ-bejaia.dz

Yazid Mohand
LaMOS Laboratory
Faculty of Exact Sciences
University of Bejaia
mohand.yazid@univ-bejaia.dz

Abstract—
The future generation of WiFi networks is expected to guarantee a best service for Real-Time Applications in IoT scenarios, namely WiFi 7 for EHT WLAN. WLANs are the direct access networks to the Internet thanks to their features. In this context, Up-Link OFDMA Random Access or UORA, originally introduced in IEEE 802.11ax standard is the notable feature that can response at best to the requirements of such applications in WiFi 7. In this paper, we aim at studying the performance of UORA by measuring the collision rate, the resource allocation rate and the Up-Link throughput. Thus, we implement and simulate the UORA algorithm under MATLAB programming language. Then we show through the results that UORA allows low throughputs, and therefore the resource allocation strategy requires some upgrades in order to effectively use the spectrum.

Keywords—
802.11be WLAN, IoT Applications, OFDMA, UORA, Performance Evaluation

I. Introduction

Internet has continued to evolve from the Internet of Contents, to the Internet of Services, then the Internet of People and now to the Internet of Things. Internet of Things (IoT) is a new paradigm that refers to an interconnection of an enormous number of devices capable of sensing and transmitting information about a specific environment or people [1]. In this context, there are various applications for IoT devices such as smart home networks, smart city networks, and smart grid systems, and there exist a variety of communication scenarios. Basically, the IoT applications adopt the 802.11 WLAN client-server communications [2], in where the WLAN is the direct access network that allows the connection of everything.

IEEE 802.11be branded WiFi 7 and known as Extremely High Throughput (EHT) is the future amendment of the IEEE 802.11ax amendment. It promises a data rates explosion of up to 46 Gbits/s, lower latency, and better network management when multiple devices are connected. The candidate features of IEEE 802.11be are mainly proposed at the physical and Medium Access control layers. Among the direct enhancement of 802.11ax, the extension of the channel bandwidth from 160 MHz in WiFi 6 to 320 MHz, the Quadrature Amplitude Modulation (QAM) scheme is quadrupled to 4096 in WiFi 7, WiFi 7 also doubles the number of simultaneous spatial streams from 8 in WiFi 6 to 16 on a wireless link Multiple Input Multiple Output or MIMO (16x16x16) [3]–[5]. For the novel candidate features, the multi-band transmission technique or Multi-Band Link Operation (MLO) which allows the simultaneous use of 2.4, 5 and 6 GHz [6], [7], the Multiple Access Points (APs) Coordination schemes including Coordinated OFDMA (Co-OFDMA) [8] and Coordinated Joint Transmission (Co-JT) or as also named Distributed MIMO (D-MIMO) [9].

Unlike the existing WiFi networks, such as 802.11a/b/g/n/ac, which allow a single user to access to the whole channel, the IEEE 802.11ax amendment introduces Multi-User (MU) orthogonal frequency-division multiple access (MU-OFDMA) on both 2.4 and 5 GHz band to improve the efficiency of the networks in dense deployments. Which corresponds exactly with the IoT ultimate goal that aims at supporting an extremely number of devices. In OFDMA, the channel is divided into sub-carriers, These sub-carriers are grouped in standard combinations. These combinations of sub-carriers or *tones* form the Resource Units (RUs) which are allocated to multiple users [10]. An hybrid access is proposed for MU-OFDMA including two modes: Scheduled Access (SA) and Random Access (RA) where the communications are organized into sequences of two cycles comprising each cycle a DownLink (DL) phase and an UpLink (UL) phase. Each access mode of MU-OFDMA consists on two phases, that are: the Resource allocation phase and the data transmission phase. In the DL OFDMA phase, the AP transmits packets simultaneously to multiple STAs using a different RU for each STA. In the UL

OFDMA phase, several STAs simultaneously transmit packets to the AP, each STA using a different RU. According to the research works in the literature, one of the notable features of the IEEE 802.11be standard is the enhancement of the UL OFDMA RA or as named UORA, this is because the IoT communications are focused on the density deployments and based client-server, i.e the communications that comes from the devices into the Access Point (AP) in order to access to the Internet and transmit their information [11]. This is why we are interested, in this paper, to firstly give the capabilities of the UORA feature.

The rest of this paper is organized as follows: Section 2 is divided into two sub-sections: Background and Related Works. We provide a comprehensive review of the key WiFi 7 features and IoT concepts, then a related works to the current status and directions given in the literature. We devote Section 3 for the performance study of the UORA technique. Finally, in section 4, we end our paper with a conclusion.

II. Background and Related works

This section is divided into two sub-sections. In the first one, we describe the WiFi 7 and IoT. In the second one, we review some research works related to WiFi 7 based IoT.

A. Background

We provide in the first part, the key WiFi 7 technologies in both Physical (PHY) and Medium Access Control (MAC) layers. In the second part, we reminder the IoT concepts.

1) WiFi 7: WiFi 7 also known as IEEE 802.11be introduces technologies that are under discussion. It aims mainly to provide wider high-bandwidth, lower latency and higher reliability, especially in the case of heavily traffics and dynamic communication networks. With IEEE 802.11be, the key PHY features are already known and include as shown in Fig.1:

- *Direct Enhancements of 802.11ax*: The enhancements are PHY. The next generation of advanced modulation scheme 4K QAM (Quadrature Amplitude Modulation), increases throughput by 20 %. 4K QAM is capable of carrying 2^{12} symbols (12 bits), compared to 1K QAM used in WiFi 6, which only carries 2^{10} symbols (10 bits). Doubling the maximum channel bandwidth available to each device to 320 MHz in the 6 GHz band effectively doubles the throughput. Morever, WiFi 7 aims to double the maximum number of supported single-user MIMO (SU-MIMO) and multi-user MIMO (MU-MIMO) spatial streams to 16, with a consequent increase in capacity. In the case of MU-MIMO, the TG agreed on limiting the maximum number of spatially multiplexed Stations and streams per STA to maximally 8 [8].

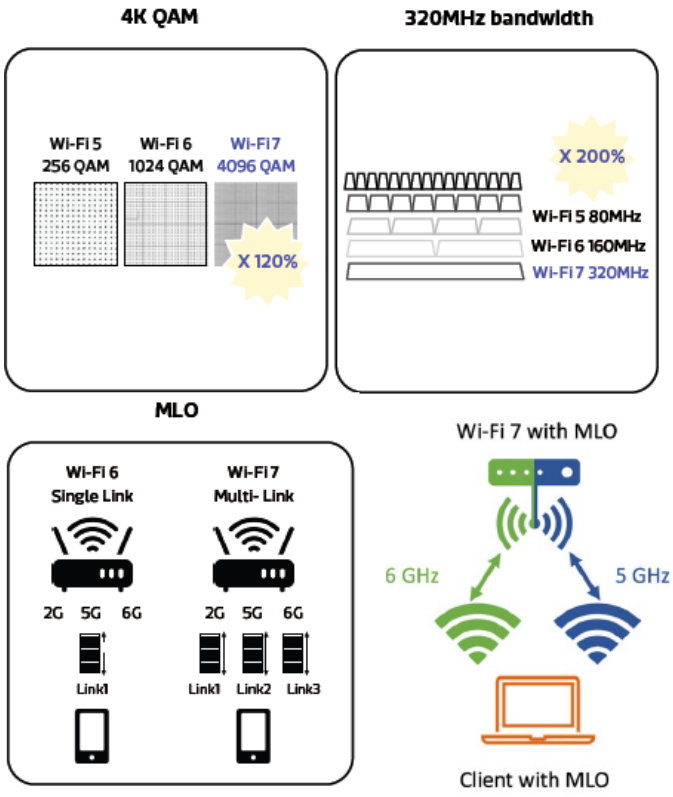


Figure 1. Key WiFi 7 PHY technologies [6], [7]

- *Multi Link Operation (MLO)* : it is a new way to utilize the three bands by increasing the users data rates. With MLO, Multi-Link Devices (MLDs) can simultaneously use the 2.4 GHz, 5 GHz and 6 GHz bands under different circumstances. The MLO feature is considered as a smart MAC-layer solution for using simultaneously multiple links. It consists on bonding multiple links (radios) in different bands and channels to work as one virtual link between the connected peers, in where, each link can work independently and simultaneously with other links.
- *Multiples RU (MRU)* : OFDMA WiFi 6 divides the radio channel into smaller frequency allocations called Resource Units. By partitioning the channel, smaller data packets can be transmitted to multiple users simultaneously, which increases throughput and reduces latency in a dense environment. WiFi 7 builds on this foundation with a new Multiple RU (MRU) feature supported in the PHY. A MRU consists of combinations of either 26, 52, 106, 242, 484, 996, 2x996, or 4x996-tone RU. The RU242, RU484, RU996, and RU2x996 correspond to the entire 20 MHz, 40 MHz, 80 MHz, and 160 MHz/80+80 MHz band, respectively. Each RU wider than RU26 can be split into two approximately twice-narrower RUs, plus one RU26 in the case of RU242 and RU996. RUs under 242-tone RU are defined as small size RUs,

while those that are equal to or larger than 242-tone RUs are defined as large size RUs, and they can only be combined with other large size RUs to form large size MRUs [6], [7].

- *Coordinated OFDMA (Co-OFDMA)* : In 802.11be, an AP is able to share its frequency resources in multiples of 20 MHz channels with a set of neighboring APs and allocates frequency resources among them. After that, both master AP and other APs send DL data or schedules UL transmission simultaneously by using OFDMA. Then, each AP occupies a part of the resource unit [8].
- *Distributed MIMO (D-MIMO)* : The distribution comes from the fact that all the used antennas must not necessarily be on a single access point. It allows APs to perform joint data transmissions to multiple STAs by reusing the same time/frequency resources. Coordinated-Joint Transmission(Co-JT) refers to a master AP that successfully accesses the wireless channel and coordinates with other APs. All antenna resources from multiple APs and all users from multiple networks on a common medium are pulled through together as shown in Fig.2. After that, both master AP and other APs send DL data to multiple STAs by using MIMO. It is noteworthy that co-JT allows AP to serve STAs and enables multiple AP to serve the same STA [9].

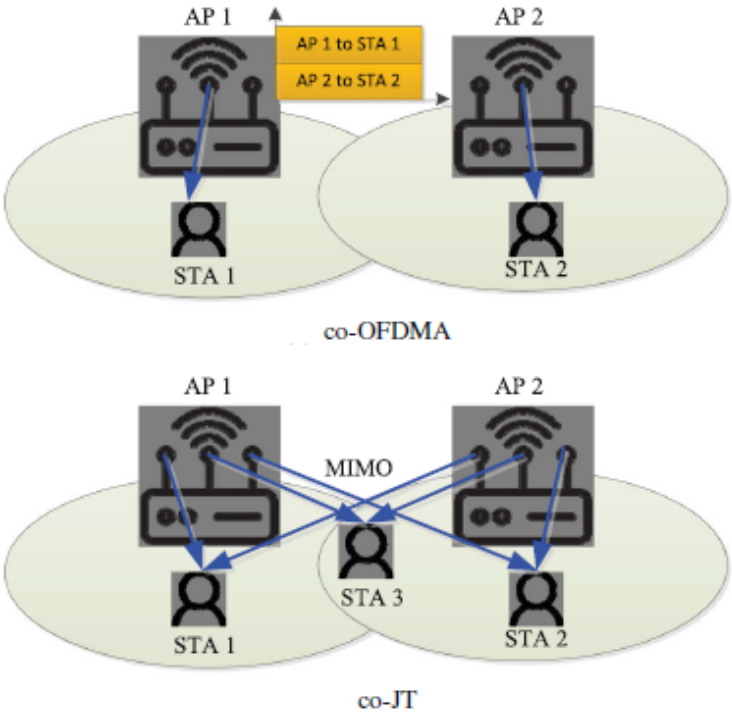


Figure 2. Co-OFDMA and Co-JT [9]

2) IoT: We remind the main concepts related to IoT as defind in [12].

- *An object Connected* is a device that can interact with the physical world independently without human intervention (it is Machine to Machine). It must be adopted for a specific use, it has some form of intelligence, an ability to receive, transmit data with software through sensor, energy, etc. There are traditional things like computers, smartphone, etc, and there are the new things like household appliances, measuring instruments, robots, locks, machine tools drones, toys, watches, vehicles, etc.
- *Internet of Things* represents a global network with self-configuration capabilities based on standards and communication protocols. The things could be physical (Hardware) or virtual (Software). IoT is a set of connected objects or things and network technologies that combine:
  - Physical things equipped with sensors.
  - Communication networks, wireless or not.
  - Remote storage spaces for saving data produced or received, stored or transmitted.
  - Data processing unit that treats data and engages the decisions processes.
- *IoT Architecture* it is divided into perception, network, support and applications layers.
  - Perception layer : it contains nanotechnologies to identify physical objects and collect the required information using integrated sensors.
  - Network layer : it performs the function of transferring the collected data to the processing system in order to read the information encoded in the data.
  - Support layer : it is the main processing unit that treats the data and makes decisions.
  - Application layer : it contains the new applications that are particularly developed to meet the needs of industry or users.
- *IoT Applications* affect all everyday life, they include health and monitoring systems to help people, industries, smart cities, homes and networks, agriculture, autonomous vehicles, etc.

B. Related works

several research works in the literature have investigated WiFi 7 based IoT Applications. Zhang et al in [13] considered WiFi 7 in the enterprise. They have proposed a novel multi-AP coordination system architecture aided by a centralized AP controller (APC) in order to decrease the collision probability of channel access. Yang et al in [14] have highlighted that it is important to cope with the collisions caused by the high number of STAs in a single coverage area. To do this, the authors have proposed a MAC scheme for AP and Multi-Band Operation (MB-Oper) coordination in compatibility with the full-duplex communications. Ahn et al in [15] have proposed a new transmission scheme which

uses the coordinated multi-AP and coordinated OFDMA concepts already introduced in 802.11be networks. The proposed transmission follows the IEEE 802.11 channel bonding rule for allowing the APs to share Transmission Opportunities (TXOPs) with other APs in coordinated APs when a new bandwidth is available again. Avdotin et al in [4] have proposed an improvement of UORA which consists on three algorithms: NUORA (Noise Resistant Uplink OFDMA Random Access), NGRA (Noise Resistant Group Resource Allocation) and NCRA (Noise Resistant Cyclic Resource Allocation). Kim et al in [16], Yang et al in [17] and Jiang et al in [18] have proposed modifications in the OFDMA functioning in case of dense deployment IoT, and Qadri et al in [10] in case of healthcare IoT (H-IoT).

The most of research works in the literature have focused on enhancing the OFDMA technique especially in the case of real-time applications. In IoT and RTA (Real-Time Applications), the communications are based on Access Points which plays the coordinator role. This is why, UpLink OFDMA Random Access or UORA is the notable feature of the future IEEE 802.11be standard, i.e the communications that comes from the devices or the objects into the Access Point (AP) in order to access to the Internet and transmit their information. Thus, we are interested to the OFDMA technique, especially, the UORA technique.

III. Performance Evaluation

This section is divided into two sub-sections. In the first subsection, we describe the UORA operating. In the second subsection, we evaluate the performance of UORA according to the simulation results.

A. UP-Link OFDMA Random Access (UORA)

To enable the MU operating with OFDMA, the AP can divide the whole channel into groups of smaller sub-channels referred as orthogonal sub-carriers, and different number of sub-carriers form a resource unit (RU). A 20 MHz channel in the 802.11ax is composed of 256 sub-carriers which are bonded in groups of 26, 52, 106 and 242 tones. However, the wide channels of 40 MHz, 80 MHz and 160 MHz channels are divided into 18, 37 and 74 RUs, respectively. Then multiple users can access and parallelly transmit over the RUs. In addition, OFDMA has been introduced for both Up-Link (UL : From Stations to AP) and Down-Link (DL: From AP to Stations) transmissions. In the UL, OFDMA allows two different types of multi-user operations, namely: Scheduled Access and Random Access, then the AP can select RUs for SA (SA-RUs) and others for RA (RA-RUs) transmissions. In what follows, we are interested to the UL OFDMA RA or as named UORA.

In the UORA mechanism, Multiple STAs can transmit data frames at the same time over different RUs. Types of RUs, their number, and their placement on a 20 MHz channel are illustrated in Fig.3. For the operation of MU transmission, UORA introduces the OFDMA Contention Window (OCW) and OFDMA Backoff (OBO) counter. To enable the UORA procedure, the AP first sends a control frame named Trigger Frame (TF). The TF contains several field of information, such as the eligible Random Access RUs (RARUs) and the corresponding Association IDentifiers (AIDs). In IEEE 802.11ax, the AID of RA-RUs is either 0 or 2045: RA-RUs have an AID equal to 0 and can be accessed by the associated STAs, whereas the unassociated STAs can occupy RA-RUs with an AID of 2045. The UORA operating principle is detailed in Algorithm 1, and summarized in the following points:

- Before enabling the UORA procedure, each STA initializes the OCW according to minimal OFDMA CW or OCWmin previously communicated by the AP using the UORA Parameter Set.
- All the STAs should wait for a TF frame. The TF frame conveys the RARUs number that are reserved for the UL transmission.
- All the STAs with a received TF has to generate a random OBO within the OCW range. Next, the STAs have to decrease their OBO by the number of RARUs. Indeed, the STAs can deduce the number of RARU by checking the AID field of the TF frame received previously from the AP.
- After updating the OBO, each STA has to check the value of the OBO counter. If the OBO counter is negative or equal to 0, the STA is allowed to randomly select an RU from the number of RARUs in order to transmit their respective data frames. Otherwise, i.e the OBO value is more than 0, the STA has to save the value of OBO and wait for the next TF frame.
- After a simultaneous transmissions through the several RUs, collisions can occur if several STAs attempt access to the same RU. In a such case, we have at least one collided RU. In consequence, all the collided STAs have to doubles their respective OCW in order to reattempt the transmission each time until the OCWmax is reached. When OCW reaches the OCWmax value, the data frame is destroyed, and the STA can compete to access the channel for a new data frame.
- After a successful transmission, the OCW is set anew to the OCWmin.

B. Simulation parameters and results

To evaluate the performance of the UORA technique, we have implemented and simulated its algorithm using MATLAB programming language. In term of performance metrics, we are interested in measuring the collision rate, the allocation rate and the UL throughput. Thus, simulations according to various parameters, see Tab.I, have been performed, in order to study the UORA performance in the following cases:

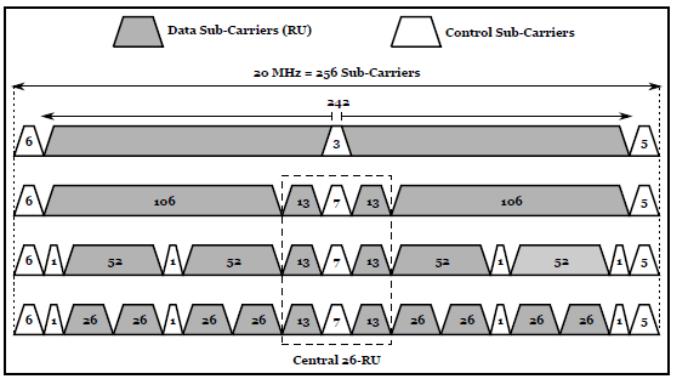


Figure 3. Resource units over a channel of 20 MHz

Algorithm 1: UORA Algorithm

: Initialize a OFDMA Contention Window $OCW = OCWmin$;
: Trigger Frame (TF) is received ;
: At all STAs with TF, Select a Random OFDMA Backoff $OBO = Random[0, OCW]$;
: At all STAs with TF, Decrease OBO $OBO = OBO - NumberRARUs$;
if $OBO(STA) <= 0$ **then**
    Assign a random RU from Number RARUs ;
    Go to (12);
else
    Do not assign and wait for the next TF ;
    Go to (4);
: Transmit Data frame over the assigned RU ;
if *Is Tx Successful* **then**
    Go to (1);
else
    Increment OCW = 2 x OCW until OCW=CWmax ;
    if *OCW = CWmax* **then**
        The Data frame is destroyed ;
        Go to (1);
    else
        Go to (2);

- Impact of RARUs number under 20 MHz channel,
- Impact of stations number under 20 MHz channel.

In what follows, we define the performance metrics:

- The collision rate represents the average collision rate during all the cycles. The collision rate during all the cycles represents a vector of size "Number of cycles", where each box i of the vector contains the collision rate during the cycle i. Thus, the average collision rate represents the total number of collisions during all cycles.
- The allocation rate represents the average resources allocation rate during all the cycles. The allocation rate during all the cycles represents a vector of size "Number of cycles", where each box i of the vector contains the RARUs allocation rate during the cycle i. Thus, the total RARUs allocation rate used during the simulation time is the average of the total RARUs allocation rates used during all cycles.
- The UL throughput represents the average quantity of Up-Link data successfully transmitted by all the accessing stations during all the cycles.

Table I. Simulation Parameters

Parameter	Description	Value
BW	Band Width	20 MHz
CW	Contention Window	[7,31]
$OCWmin$	OFDMA Minimal CW	7
$OCWmax$	OFDMA Maximal CW	31
OBO	OFDMA BackOff	Random[0,31]
TF	Trigger Frame length	at least 28 Bits
RU	Resource Unit	26 tones
$NB - RARU$	Random Access RU Number	[1,9]
$NB - Cycles$	Cycles Number	1000
$NB - STAs$	Stations Number	[4,40]
Nss	Number of Spatial Streams	1
GI	Guard Interval	0.8 μs

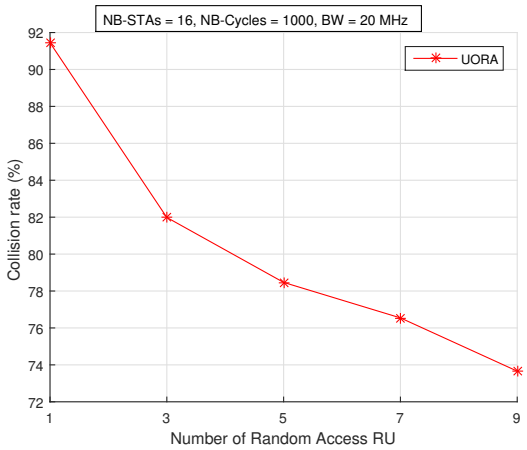


Figure 4. Collision rate Vs Number of Random Access RU

In Fig.4 and Fig.5, we study the variations of the collision rates according to respectively the available RARUs number and the stations number when executing UORA method. In Fig.4, we note that the collision rate decreases proportionally and logically with the increase of the number of RARUs. For a fixed number of stations to 16 and one single RU, the collision rate reached the peak of approximately 92 % , this is because all the stations of the network compete for accessing to one RU. After that, we remark that the rate decreases slightly with the increase of the RARUs number. That is due to the fact that, more the RARUs number increases, more the competing stations number for one RU decreases, more the RARUs number successfully used or again won increases, more the collision rate decreases. However the

rate values still remain high for all the variations of RARUs number.

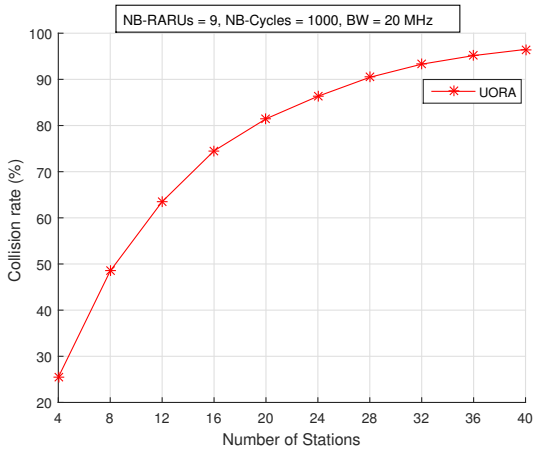


Figure 5. Collision rate Vs Number of Stations

In Fig.5, we fix the number of RARUs to 9 and vary the number of stations from 4 to 40. We show that, the collision rates are high for all the values of the stations number. Unfortunately, even for 4 stations and 9 available RARUs, i.e the number of stations is less than the number of RARUs, the collision rate reached approximately 25% which clearly highlight the inefficiency allocation RU strategy of UORA method. For other values that are more greater, we see that the collision rate increases with the increase of the stations number. In fact, the larger the network size, the higher the probability of collision.

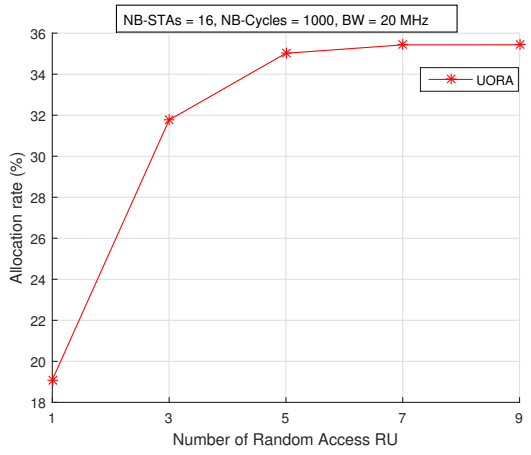


Figure 6. Allocation rate Vs Number of Random Access RU

In Fig.6 and Fig.7, we study the impact of respectively the RARUs number and the stations number on the obtained RARUs allocation rate when executing UORA. In Fig.6, with a fixed number of stations to 16, the allocation rate increase with the increase of RARUs number. With one single RARU, the obtained allocation rate is logically

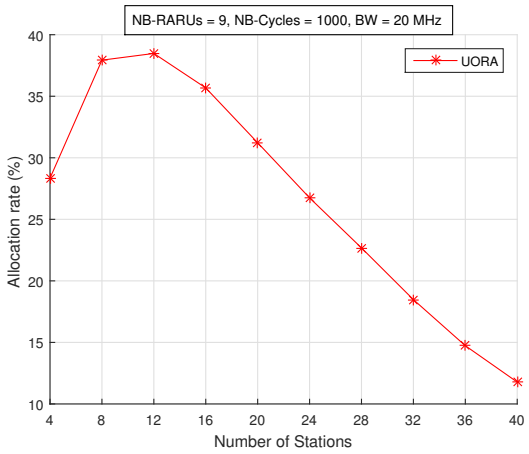


Figure 7. Allocation rate Vs Number of Stations

low and equal to approximately 19 %, besides at this point, we have obtained the higher value of collision rate. In fact, the greater the number of RARUs, the higher the probability of accessing to RUs, the higher the allocation rate.

In Fig. 7, with a fixed number of RARUs of 9, we distinguish two cases, the case where the stations number is lower than the RARUs number which is equal to 9. And the other case where the stations number is greater than 9. In the first case, it is impossible to occupy all the available 9 RARUs even we got the best case of RUs allocations (i.e one RU is assigned to exactly one station). That is why we have marked an increase of the allocation rate when passing from 4 stations to 8 stations then to 12. In the second case, i.e the stations number is greater than the RARUs number, the allocation rate decreases with the increase of the stations number. In fact, the higher the number of stations, the lower the probability of accessing to RUs, the lower the allocation rate.

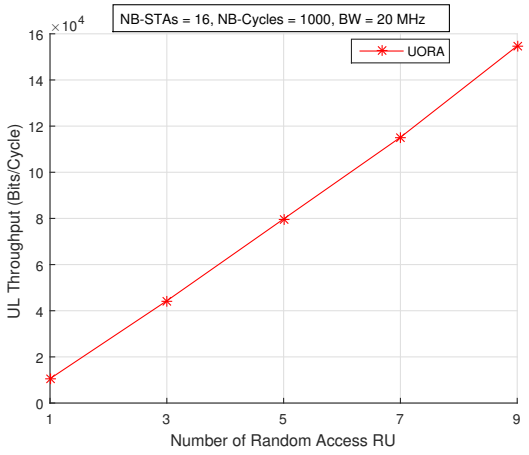


Figure 8. UL Throughput Vs Number of Random Access RU

In Fig. 8 and Fig.9, we study the variations of

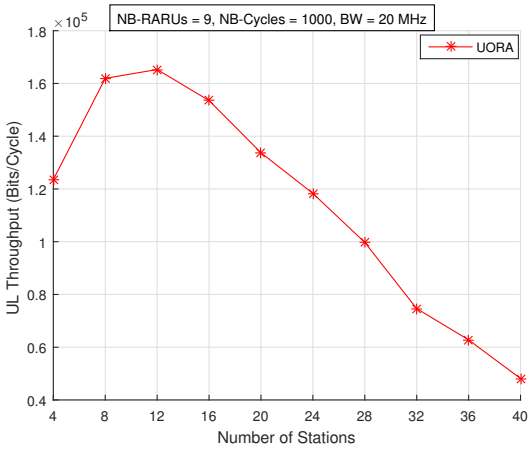


Figure 9. UL Throughput Vs Number of Stations

UL throughput according to respectively the available RARUs number and the stations number when executing UORA method. In fig.8, the UL throughput increases proportionally and logically with the increase of the number of RARUs. The greater the number of RARUs, the higher the allocation rate, the higher the UL throughput. In fig.9, for 9 RARUs, and stations number lower than 9, the throughput increases with the increase of the number of stations. However, from 12 stations, we note a proportional decrease of throughput with the increase of the stations number. Indeed, the greater the number of stations, the higher the probability of collision, the lower the UL throughput.

In summary, with the UORA random allocation of RARUs, we can obtain a certain variations in results even for the same parameters values and even if the stations number is lower than the RUs number. In the best case, i.e each STA wins one RU, we obtain 9 used RUs, 0 collided RUs and 0 non-used RUs. However, in the worst case, i.e all the stations compete for one RU among the 9 RUs, we obtain 1 collided RU, 8 non-used RUs and 0 used RUs.

IV. Conclusion

In this paper, we have been interested to the WiFi 7 based IoT Real-Time Applications. We have studied the UORA feature and addressed the critical issues of efficient resource allocation and collision because it is for stations to select the resource unit to further transmit the data. This is why, there is three states for a single RU : successful, idle or non-used and collided. In addition, when a STA fails to transmit the data frame after several attempts, it has to resume another transmission in the next cycle, which increases both delay transmission and latency.The study has been done by calculating three parameters, that are: the collision rate, the allocation rate and the Up-Link throughput and by varying the parameters of the number of the Random Access RUs (RARUs) and the number of stations. Unfortunately, one of the UORA problems is the high probability of collision due to the high number of simultaneously accessing stations to RUs. A successful RU means that only one station chooses this RU at one time and a such case is almost impossible with the UORA allocation strategy based random.

References

[1] J.O Agyemang, and Kponyo, J Klogo, G. S Boateng and J. Boateng, "Lightweight rogue access point detection algorithm for WiFi-enabled Internet of Things (IoT) devices, *IEEE/ACM transactions on networking*, 2018, 26(2), pp 864–878.

[2] I. Md Manowarul, F. Nobuo and Sudibyo, R. W Munene, K. I Kao and W. C Kao, 'A dynamic access-point transmission power minimization method using PI feedback control in elastic WLAN system for IoT applications," *Internet of Things*, 2019, 8, pp 100089.

[3] E. Au, 'IEEE 802.11 be: Extremely high throughput, *IEEE Vehicular Technology Magazine*, 2019, pp 138-140.

[4] E. Avdotin, D. Bankov, E. Khorov, and A. Lyakhov, 'Resource allocation strategies for real-time applications in Wi-Fi 7, *2020 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, 2020, pp 1-6.

[5] R. P. F Hoefel, 'IEEE 802.11 be: throughput and reliability enhancements for next Generation Wi-Fi networks, *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, 2020, pp 1-7.

[6] MediaTek, Key advantages of WiFi 7,Performance, MRU and MLO, *White Paper*, 2022.

[7] M. HOSSAM, A. KAMEL, The Future of Smart City WiFi How Wi-Fi 7 Enhances the Way of, *CWNP CWNE Candidate Whitepaper Series*, 2023.

[8] G.A. Adrian, L.P. David, G.G. Lorenzo, and G. Giovanni, IEEE 802.11 be: Wi-Fi 7 strikes back, 59(5), *IEEE Communications Magazine*, 2022, pp 102-108.

[9] M. Yang, Mao and B. Li, Survey and perspective on extremely high throughput (EHT) WLAN—IEEE 802.11 be, 25, *Mobile Networks and Applications*, 2020, pp 1765–1780.

[10] Y.A Qadri, N. Zulqarnain, M. Ali, G.V Arslan , Eduard and K. S Won, Preparing wi-fi 7 for healthcare internet-of-things, 22(16), " *Sensors*, 2022, pp 6209.

[11] E. Khorov, Evgenyn, I. Levitsky, I. F. Akyildiz, 'Current status and directions of IEEE 802.11 be, the future Wi-Fi 7, 8, *IEEE Access*, 2020, pp 88664-88688.

[12] I. Saleh, Carey, Internet des Objets (IdO): Concepts, enjeux, défis et perspectives, *Revue Internet des objets*, 2018.

[13] L. Zhang, H. Yin, S. Roy, and L. Cao, "ultiaccess point coordination for next-gen Wi-Fi networks aided by deep reinforcement learning, *IEEE Systems Journal*, 2022, 17(1), pp 904–915.

[14] M. Yang, B. Li, Z. Yan, Y. Yan, "AP coordination and full-duplex enabled multi-band operation for the next generation WLAN: IEEE 802.11 be (EHT), *2019 11th international conference on wireless communications and signal processing (WCSP)*, 2019, pp 1-7.

[15] W. Ahn, Novel multi-AP coordinated transmission scheme for 7th generation WLAN 802.11 be, *Entropy*, 2020, 22(12), pp 1426.

[16] Y Kim, L Kwon, EC Park, OFDMA backoff control scheme for improving channel efficiency in the dynamic network environment of IEEE 802.11 ax WLANs, *Sensors*, 2021, 21(15), pp 5111.

[17] A Yang, B Li, M Yang, Z Yan, Y Xie,Utility optimization of grouping-based uplink OFDMA random access for the next generation WLANs, *Wireless Networks*, 2021, 27, pp 809-823.

[18] Z Jiang, B Li, M Yang, Z Yan, Latency oriented OFDMA random access scheme for the next GENERATION WLAN: IEEE 802.11 be, *Smart Grid and Internet of Things: 4th EAI International Conference, SGIoT 2020, TaiChung, Taiwan, December 5–6, 2020, Proceedings*, 2021, pp 351-362.

Analyse de la fiabilité et des coûts du modèle $M^{[X]}/G_1,G_2/1$ avec rappels et pannes du serveur

Zirem Djamila
Unité de Recherche LaMOS, Université de Béjaia, 06000 Bejaia, Algerié
ziremdjamila@yahoo.fr

Boualem Mohamed
Unité de Recherche LaMOS, Université de Béjaia, 06000 Bejaia, Algerié
mohammed.boualem@univ-bejaia.dz

Aïssani Djamil
Unité de Recherche LaMOS, Université de Béjaia, 06000 Bejaia, Algerié
djamil.aissani@univ-bejaia.dz

Résumé—

Ce travail s'intéresse à l'étude des indices de fiabilité et à l'analyse des coûts du modèle d'attente $M^{[X]}/G_1,G_2/1$ avec rappels, deux phases service, pannes et réparations du serveur et clients impatients. Le modèle en question peut être utilisé pour la modélisation des systèmes de contrôle de la production dans le domaine économique et les systèmes de gestion des stocks stochastiques. Par exemple, la production à la commande est une politique de production en planification et en gestion de production. On suppose qu'un processus de production, où la machine produit certain articles, peut nécessiter deux phases de services, la première est le traitement des matières premières et la seconde consiste à la vérification. Des pannes surviennent à n'importe quel moment, et qui sont réparés instantanément.

Mots-Clés—

File d'attente avec rappels, deux phases service, pannes et réparations, fiabilité, coût du modèle, impatience.

I. INTRODUCTION

Les mesures de fiabilité, en collectant et en évaluant des données pertinentes, offrent un aperçu précieux permettant d'identifier les domaines qui requièrent des ajustements pour améliorer le système en question [1].

En revanche, il convient de souligner que l'analyse des coûts demeure l'élément prédominant dans toutes les situations pratiques, à chaque phase du processus. La conception optimale d'un système de files d'attente avec rappels implique la détermination des paramètres optimaux du système, tels que le taux de service moyen optimal, le nombre optimal de serveurs, et le taux d'attente moyen optimal, en utilisant des fonctions de coût spécifiques. Cela s'explique par le fait que la principale préoccupation de tout gestionnaire est d'optimiser le coût moyen total du système [2].

Notre travail se concentre, dans un premier temps, sur l'étude de la fiabilité du système d'attente $M^{[X]}/G_1,G_2/1$ avec rappels, deux phases service, pannes et réparations du serveur et clients impatients dans le but de pouvoir déterminer les indices de fiabilité essentiels, à savoir: La probabilité de la disponibilité du serveur (Availability), la fréquence de la défaillance (Failure frequency) et la fréquence de la défaillance en régime stationnaire du serveur. Dans un dernier temps, l'étude est focalisée sur l'analyse des coûts du modèle considéré.

II. DESCRIPTION DU MODELE

Nous considérons un nouveau modèle de files d'attente $M^{[X]}/G_1,G_2/1$ avec un seul serveur et clients impatients, le serveur est sujet à des pannes actives et deux phases de services. Les clients arrivent dans le système par groupes selon un processus de Poisson de taux λ. A l'arrivée d'un groupe de clients, si le serveur est occupé ou en panne, dans ce cas le groupe entre en orbite avec une probabilité p, sinon il quitte le système sans être servi avec une probabilité $1-p$. Par contre, si le serveur est libre, le premier client qui est en tête du groupe commence son service et le reste du groupe entre en orbite suivant une discipline FCFS pour tenter à nouveau. Lorsque le serveur est libre, le client en tête dans l'orbite est en compétition avec le client primaire.
Si le client primaire arrive le premier par rapport au client en tête de l'orbite, ce dernier retourne en orbite avec une probabilité q, sinon il quitte le système sans être servi avec une probabilité $1-q$. On considère deux phases de service, la première phase essentielle FES (first Essential Service) qui est obligatoire pour tous les clients. Après avoir terminé le *FES*, le client peut demander la deuxième phase de service *SOS* (Second Optionnel Service), qui est facultative avec une probabilité b ou quitter le système avec une probabilité $1-b$ sans demander le service SOS.

Les clients dont le service est interrompu décident de rester devant le serveur avec une probabilité r_i, ou bien de rejoindre l'orbite service avec une probabilité $1-r_i$ avec $i=1,2$. Une fois la réparation achevée, le serveur reprend le service du client. Le serveur n'est pas autorisé à accepter de nouveaux clients jusqu'à ce que le client en service quitte le système [2-9]. Fig. 1. illustre le modèle considéré.

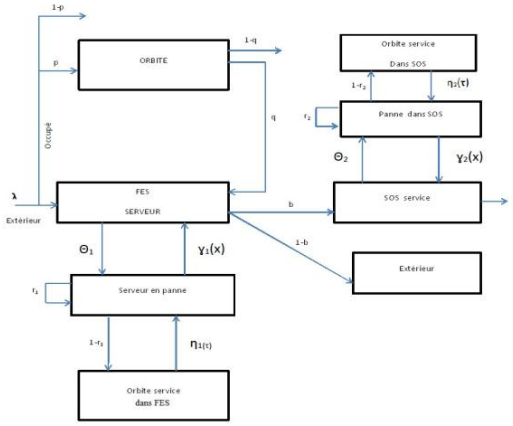


Figure 1. Schéma représentatif du modèle étudié

III. MESURES DE FIABILITE

Dans cette section, nous nous penchons sur l'examen de divers indices de fiabilité relatifs au système que nous considérons. Plus précisément, nous entreprenons une analyse approfondie de la disponibilité du serveur et de la fréquence de défaillance associée à ce dernier. Il est essentiel de comprendre que les pannes du serveur peuvent avoir des répercussions significatives sur la qualité du service dans pratiquement tous les domaines, que ce soit dans le cadre des systèmes informatiques et de communication, des processus de fabrication ou de production.

Dans le contexte d'un système de file d'attente avec rappels, où le serveur est sujet à des défaillances, les mesures de fiabilité se révèlent cruciales pour fournir les informations essentielles à l'amélioration continue du système [1,9]. Elles permettent de déterminer les domaines vulnérables nécessitant des ajustements pour renforcer la fiabilité, réduire les temps d'arrêt et améliorer l'expérience globale de service pour les utilisateurs.

La probabilité de la disponibilité A_v (Availability) et la fréquence de défaillance F_f (Failure frequency) du serveur sont respectivement donnés :

- La probabilité de la disponibilité A_v (Availability) :

$$A_v = P_{00} + \lim_{z\to 1}[P_0(z) + P_1(z)]$$

- La fréquence de défaillance F_f (Failure frequency) du serveur :

$$F_f = \mu \lim_{z\to 1}[P_1(z)]$$

Théorème :

✓ La disponibilité du serveur (A_v) , qui est la probabilité que le serveur soit disponible à l'instant t :

$$A_v = \frac{(1-q+qL_A(\lambda))(1+B_1) - L_A(\lambda)(K'(1)) + \lambda B_1(1-g'(1))}{[(1-q+(q-pg')L_A(\lambda))\left(1+\frac{K'(1)}{pg'(1)}\right) + L_A(\lambda)(pg'(1)-K'(1)(\frac{1}{pg'(1)}-\frac{1}{p}))]}$$

✓ La fréquence de la défaillance en régime stationnaire du serveur correspond à la probabilité de panne du serveur en cours de service :

$$F_f = \frac{\lambda\mu B_1\{-(1-q+qL_A(\lambda)) + L_A(\lambda)(1-g'(1))\}}{[(1-q+(q-pg')L_A(\lambda))\left(1+\frac{K'(1)}{pg'(1)}\right) + L_A(\lambda)(pg'(1)-K'(1)(\frac{1}{pg'(1)}-\frac{1}{p}))]}$$

IV. ANALYSE DES COUTS DU MODELE

L'objectif de cette analyse est de minimiser le coût total, qui équivaut à la somme de deux coûts : le coût associé à la capacité de service mise en place (coût de service) et le coût associé à l'attente des clients (coût d'attente). Le coût de service est le coût résultant du maintien d'un certain niveau de service. Par exemple, le coût associé au nombre de caisses dans un supermarché, au nombre de réparateurs dans un centre de maintenance, au nombre de guichets dans une banque, au nombre de voies d'une autoroute, etc. En cas de ressources inoccupées, la capacité est une valeur perdue, car elle est non stockable. Les coûts d'attente sont constitués des salaires payés aux employés qui attendent pour effectuer leur travail (mécanicien qui attend un outil, chauffeur qui attend le déchargement du camion, etc.), du coût de l'espace disponible pour l'attente (grandeur de la salle d'attente dans une clinique, longueur d'un portique de lave-auto, kérosène consommé par les avions qui attendent pour atterrir) et, bien sûr du coût associé à la perte de clients impatients qui vont chez les concurrents. En pratique, lorsque le client est externe à l'entreprise, le coût d'attente est difficile à évaluer, car il s'agit d'un impact plutôt qu'un coût pouvant être

comptabilisé. Cependant, on peut considérer le temps d'attente comme un critère de mesure du niveau de service. Le gestionnaire décide du temps d'attente acceptable (tolérable), et il met en place la capacité susceptible de fournir ce niveau de service. Lorsque le client est interne à l'entreprise, les clients sont des machines et des camions, l'équipe d'entretien, on peut établir directement certains coûts se rapportant au temps d'attente des clients (machines). Par ailleurs, il ne faut pas conclure trop rapidement que pour l'entreprise, le coût du temps d'attente d'un employé qui attend est égal à son salaire durant le temps d'attente, cela impliquerait que la baisse nette des gains de l'entreprise, du fait de l'inactivité d'un employé, est égale au salaire de ce dernier, ce qui, a priori, n'est pas évident. L'employé, qu'il travaille ou qu'il attende, reçoit le même salaire. Par contre, sa contribution aux gains de l'entreprise est réellement perdue, car la productivité baisse. Quand un opérateur de machine est inactif parce qu'il attend, sa force productive (qui peut comprendre, outre son salaire, une proportion des coûts fixes de l'entreprise) est perdue. En d'autres termes, il faut tenir compte non pas de la ressource physique en attente, mais plutôt de la valeur (coût) de toutes les ressources économiques inactives, et évaluer ensuite la perte de profit à partir de la perte de productivité. L'objectif de l'analyse des files d'attente est de trouver un compromis entre le coût associé à la capacité de service et le coût d'attente des clients. La figure suivante illustre bien ce concept. Notons que lorsque la capacité de service augmente, le coût de service augmente. Par souci de simplicité, nous avons illustré un coût de service linéaire. Cela n'affecte en rien la démonstration. Lorsque la capacité de service augmente, le nombre de clients en attente et le temps d'attente tendent à diminuer, donc les coûts d'attente diminuent.

Le coût total (la somme des coûts de service et d'attente. Pour cela, il suffit de déterminer le niveau de service se traduisant par le coût total minimum qu'est représenté sur le graphique par la courbe en forme de *U*. Graphiquement, il suffit de déterminer le niveau de service se traduisant par le coût total minimum (contrairement au modèle de la quantité économique utilisé dans la gestion des stocks. Dans le cas d'une clientèle externe à l'entreprise, les files d'attente donnent une image négative de la qualité du service offert. Dans cette situation, les entreprises auront tendance à augmenter la rapidité du service plutôt que d'augmenter le nombre d'employés.

.

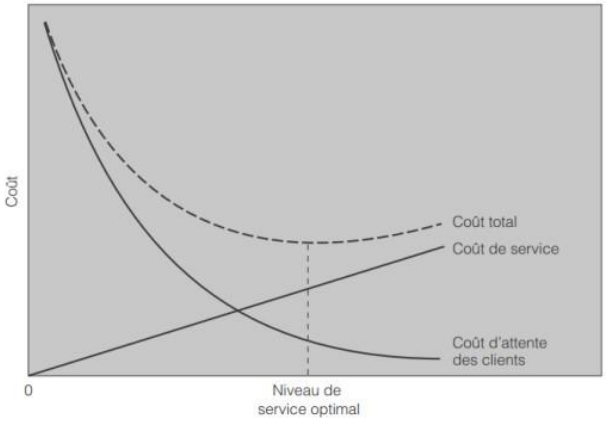


Figure 2. graphique du coût

- le coût associé à la capacité de service mise en place (coût de service).

- le coût associé à l'attente (coût d'attente).

Le coût du modèle de files d'attente étudiée est le coût total par unité de temps qui est donné par la formule suivante :

$$TC = C_h E(N_q) + C_0(1 - P_{00}) + C_s \lambda E(X) + C_a P_{00}$$

$$TC = C_h E(N_q) + C_0 \frac{E(T_b)}{E(T_c)} + C_s \frac{1}{E(T_c)} + C_a \frac{E(T_0)}{E(T_c)}$$

Avec :

C_s: Coût de démarrage par cycle occupé,

C_h: Coût du nombre de clients dans la file d'attente par unité de temps,

C_o: Coût d'exploitation par unité de temps (coût par unité de temps pour maintenir le serveur en fonctionnement),

C_a: Coût de démarrage par unité de temps (le travail préparatoire du serveur avant de commencer le service).

$E(T_0)$: est la période moyenne d'inoccupation.

$E(T_b)$ et $E(T_c)$: sont les périodes moyennes d'occupation du système et d'occupation d'un cycle qui sont données respectivement par :

$$E(T_o) = \frac{1}{\lambda E(X)},$$

$$E(T_b) = \frac{1}{\lambda E(X)}\left[\frac{1}{P_{00}} - 1\right],$$

$$E(T_c) = \frac{1}{\lambda E(X)}\left[\frac{1}{P_{00}}\right].$$

V. ILLUSTRATION NUMERIQUE

Afin d'illustrer la fonction coût ***TC***, pour les différentes valeurs des paramètres du modèle considéré numériquement, nous considérons les paramètres suivants :

$C_h = 5, C_s = 1000, C_o = 100, C_a = 100$, et en donnant les valeurs appropriées aux paramètres suivants $(\lambda, \mu, q, p, g_1, g_2, \eta_1, \eta_2, \beta_1, \beta_2, \gamma_1, \gamma_2)$

Pour les Tableaux de 1 à 4, les valeurs choisies pour satisfaire la condition de stabilité du modèle sont les suivantes :
$g_1 = 1, g_2 = 2, \eta_1 = 1, \eta_2 = 2, \beta_1 = 1, \beta_2 = 2, \gamma_1 = 1, \gamma_2 = 2$

Table 1. L'effet du taux d'arrivée λ sur le coût total (TC)

λ	*P=0.5*			*P=0.7*		
	μ = 0.01	μ = 1	μ = 1.3	μ = 0.01	μ = 0.1	μ = 1.3
0.001	105.9953	105.9916	124.7250	105.9967	105.9938	105.9906
0.026	129.8668	129.7649	129.6488	129.9050	129.8249	129.7347
0.052	154.6993	154.4760	154.2206	154.7847	154.6057	154.4009
0.076	179.4962	179.1291	178.7079	179.6371	179.3375	178.9890
1.001	204.2560	203.7206	203.1032	204.4602	204.0151	203.4888

D'après la Table 1, nous constatons que :
L'augmentation du taux d'arrivée de clients a un effet direct sur le coût total du système, et cela se traduit par un impact positif sur l'économie dans l'ensemble. En d'autres termes, lorsque le flux de clients s'accroît, le coût total du système augmente également. Cette corrélation peut être considérée comme favorable du point de vue économique, car elle reflète une plus grande utilisation des ressources, ce qui peut stimuler l'activité économique globale. L'investissement dans le système, en réponse à une augmentation du taux d'arrivée de clients, contribue ainsi à la croissance économique en créant de la demande pour les biens et services associés.

Cependant, il est important de noter que, malgré cette tendance positive, nous avons établi que le coût total minimum du système demeure à ***105.9916*** pour les paramètres optimaux. Cela signifie que, même avec une augmentation du taux d'arrivée de clients, en respectant ces paramètres optimaux, le coût total du système reste à un niveau bas de ***105.9916***. Cela suggère que les ajustements apportés au système en fonction de ces paramètres permettent d'optimiser les coûts, même en présence d'une augmentation du taux d'arrivée de clients. Cette efficacité dans la gestion des ressources contribue positivement à la stabilité économique, en garantissant un coût minimum malgré les fluctuations de la demande.

Table 2. L'effet du taux d'arrivée μ sur le coût total (TC)

μ	*P=0.5*			*P=0.7*		
	λ = 0.1	λ = 0.15	λ = 0.2	λ = 0.1	λ = 0.15	λ = 0.2
0.002	204.3005	253.7785	303.1168	20.4.4595	254.0602	303.5481
0.052	204.1897	253.5802	302.8009	204.3688	253.8947	303.2788
0.102	204.0685	253.3615	302.4492	204.2692	253.7106	302.9750
0.0152	203.9415	253.1306	302.0744	204.1644	253.5144	302.6468
0.202	203.8088	252.8876	301.6766	204.0543	253.3058	302.2933

D'après la Table 2, on constate que lorsque le taux de panne augmente, on observe une diminution du coût total du système. Cette situation peut sembler contre-intuitive, mais elle s'explique par le fait que le serveur, en raison de ses défaillances accrues, n'est plus en mesure de fournir des services aux clients de manière efficace. Par conséquent, le nombre de clients présents dans le système augmente, et la période de service devient plus longue, ce qui entraîne une perte nette de clients impatients, incapables de supporter de longues attentes. Il est évident que ces pannes ont un impact nettement négatif sur l'expérience des clients et la qualité du service dans le système.

De plus, malgré cette tendance, il est à noter que le coût total minimum demeure à ***203.8088*** pour les paramètres optimaux. Cela signifie que même en présence d'une augmentation du taux de panne, en respectant ces paramètres optimaux, le coût total du système reste relativement bas à 203.8088. Cette constatation met en évidence l'efficacité de ces paramètres dans la gestion des ressources du système, malgré les obstacles liés aux défaillances du serveur. En fin de compte, elle contribue positivement à la stabilité économique en garantissant un coût minimum, quelles que soient les fluctuations des taux de panne.

Table 3. L'effet du taux d'arrivée **p** sur le coût total (***TC***)

	μ =0.5			μ =0.7		
	λ = 0.1	λ = 0.15	λ = 0.2	λ = 0.1	λ = 0.15	λ = 0.2
0.01	153.7901	202.2358	231.1467	153.3909	201.4276	230.0915
0.25	153.8102	202.2745	231.1940	153.4157	201.4700	230.1385
0.5	153.8312	202.3151	231.2436	153.4418	201.5146	230.1884
0.75	153.8522	202.3577	231.2944	153.4680	201.5597	230.2391
1	153.8733	202.3674	231.3454	153.4943	201.6055	230.2908

En ce qui concerne la Table 3, une tendance significative se dessine : on constate une augmentation du coût total du système parallèlement à la croissance de la probabilité d'entrée des clients en orbite. Cette observation est particulièrement intéressante, car elle met en évidence une corrélation marquée entre ces deux variables. Cette augmentation du coût total du

système est principalement due au fait qu'une probabilité plus élevée d'entrée des clients en orbite entraîne une augmentation du nombre de clients servis. Plus de clients sont traités par le système, ce qui peut sembler entraîner des coûts plus élevés. Cependant, ce phénomène reflète en réalité une influence positive. En effet, une augmentation du nombre de clients servis est généralement synonyme d'une utilisation plus intensive du système, ce qui peut se traduire par une rentabilité accrue. Une forte demande et une utilisation optimale des ressources du système peuvent contribuer à une meilleure efficacité opérationnelle et à des revenus plus élevés. Par conséquent, bien que le coût total puisse augmenter, cette croissance peut être considérée comme une réaction positive à une demande plus importante. Elle indique que le système est capable de gérer une charge de travail accrue, ce qui est bénéfique d'un point de vue économique.

Table 4. L'effet du taux d'arrivée **q** sur le coût total (***TC***)

q	μ=0.5			μ=0.7		
	λ = 0.05	λ = 0.1	λ = 0.13	λ = 0.05	λ = 0.1	λ = 0.13
0.001	153.8956	202.4454	231.4166	153.5166	201.6559	230.3690
0.025	153.8902	202.4338	231.1940	153.4157	201.4700	230.1385
0.05	153.8846	202.4217	231.3815	153.5056	201.6310	230.3305
0.075	153.8790	202.4095	231.3635	153.5000	201.6183	230.3107
0.1	153.8733	202.3974	231.3454	153.4943	201.6055	230.2908

D'après la Table 4, on remarque qu'à mesure que la probabilité de rappel des clients en orbite *(q)* augmente, le coût total du système diminue. Cette tendance peut sembler contre-intuitive, mais elle trouve son explication dans le fait que le serveur, lorsque la probabilité de rappel des clients en orbite est élevée, ne peut plus fournir de service efficacement.
En conséquence, le nombre de clients en attente dans le système augmente, et les périodes d'attente se prolongent. Cette situation peut induire une perte nette de clients, car certains deviennent impatients et quittent le système. Par conséquent, le coût total du système diminue dans ce scénario, bien que cela puisse sembler paradoxal. Il est également pertinent de noter que, dans nos exemples numériques précédents, les résultats montrent que les paramètres ont une influence directe sur le coût total (***TC***) du système. Cette corrélation entre les paramètres et le coût total reflète fidèlement la réalité et souligne l'importance d'analyser minutieusement ces paramètres pour comprendre leur impact sur l'efficacité opérationnelle et l'économie globale du système. Ces constatations renforcent l'importance de la gestion et de l'optimisation de ces paramètres pour garantir des performances économiques optimales.

VI. CONCLUSION

Dans travail, nous avons effectué une analyse de divers indices de fiabilité relatifs au système considéré dans notre étude. De plus, nous avons étudié le coût du modèle d'attente $M^{[X]}/G/1$ avec rappels, pannes, réparation, clients impatient. Nous avons obtenu la formule explicite du coût et nous avons montré l'influence des paramètres du modèle considéré sur le coût total.

Pour nos exemples numériques précédents, on constate que l'influence des paramètres sur le coût total ***TC*** coïncide est reflète la réalité, cela est dû au fait que le serveur ne peut plus fournir de services aux clients, par conséquent le nombre de clients dans le système augmente et la période de service est plus longue, ce qui entraine une perte du nombre de clients dans le système (clients impatients) en raison de l'attente prolongée. Il est évident que la panne a un impact négatif sur l'économie, ce qui provoque donc d'énormes pertes.

REFERENCES

[1] D. Aïssani and A. Aïssani,"Compte rendu de la journée d'études: Modèles de Fiabilité et Sciences de l'ingénieu," *in Proceeding of Modèles de Fiabilité et Sciences de l'Ingénieur*, Béjaia, 1988.

[2] A. Aissani, "Optional control of an *M/G/1* retrial queue with vacation," *Journal of Systems Science and Systems Engineering*, 2008, 17, pp 487--502.

[3] J. R. Artalejo, I. Atencia, "On the single server retrial queue with batch arrivals, " *Indian Journal of Statistics*, 2004, 66, pp 140--158.

[4] M. Boualem, N. Djellab and D. Aissani, "Stochastic inequalities for *M/G/1* retrial queues with vacation and constant retrial policy," *Mathematical and Computer Modeling*, 2009, 50, pp 207--212.

[5] M. Boualem, N. Djellab and D. Aissani, "Approche régénérative de la file d'attente *M/G/1* avec rappels classiques et vacances exhaustives du serveur", *Journal Européen des Systhèmes Automatisés*, 2011, 45, pp 253--267.

[6] M. Boualem, N. Djellab and D. Aissani, "Stochastic approximations and monotonicity of a single server feedback retrial queue," *Mathematical problems in Engineering*, 2012, 2012, pp 1--13.

[7] G. Choudhury, L. Tadj, "An *M/G/1* queue with two phases of service subject to the server breakdown and delayed repair," *Applied Mathematical Modelling*, 2009, 33, pp 2699--2709.

[8] D. Zirem, M. Boualem, and D. Aissani, "M^X/G/1 queueing system with breakdowns and repairs", *International Journal of Advances in Computer Science & Its Applications*, 2016, 6, pp 18-21

D. Zirem, M. Boualem, K. Adel-Aissanou and D. Aïssani, "Analysis of a single server batch arrival unreliable queue with balking and general retrial time", *Qual. Technol. Quant. Manag*., 2019, 16(6), pp 672–695.

Index des Auteurs

Editions LaMOS 2023
ISBN : 978-9931-884-16-3

MFSI 2023

Il y a 35 ans de cela, la Conférence Nationale MFSI'1988 (Modèles de Fiabilité et Sciences de l'Ingénieur) était organisée à Béjaia. Il s'agissait de la première manifestation jamais organisée dans notre pays sur la fiabilité [voir le Compte Rendu publié par la revue Es-Syana – la Maintenance, Vol. 1, INMA, Ministère de l'Industrie Lourde édition, 1988, pp. 27 – 28]. Dix ans plus tard, le Ministère de la Défense Nationale, à travers l'ENITA (aujourd'hui l'Ecole Militaire Polytechnique) avait pris le relais pour l'organisation, à Bordj-el-Bahri, de la 2ème édition [voir le Compte Rendu dans la revue MATAPLI de la SMAI (Société Française des Mathématiques Appliquées et Industrielles), n° 54, 1998, pp. 65 – 66].

La 3ème édition de la Conférence Nationale MFSI aura lieu de nouveau à Béjaia du 19 au 20 novembre 2023, en tenant compte des évolutions scientifiques et technologiques de ces dernières années. Cette manifestation sera couplée avec l'organisation de la 6ème édition de la « Journée Nationale de la Fiabilité », créée en 2013 par l'Ecole Nationale Polytechnique (El Harrach), l'AD – ENP (association des diplômés), en collaboration avec l'Unité de Recherche LaMOS Béjaia (Modélisation et Optimisation des Systèmes), à la mémoire du Professeur Abdelaziz Ouabdesselem (considéré comme étant le « père » des fiabilistes algériens).

Ce livre de 185 pages regroupe les textes des 22 communications retenues par le Comité Scientifique. Ils sont répartis en 04 thèmes généraux : « Sûreté de Fonctionnement et RBDO », « Fiabilité dans les Sciences de l'Ingénieur », « Modèles Stochastiques de Fiabilité »et « Fiabilité dans les Réseaux de Télécommunication».

9 789931 884163