الجـــــــــــــــــــمهـوريـــــة الجـــــــــــــزائـريـة الديمـــــــقـراطيـة الشـــــــــعبيـة

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التـــــــــــــــــــــــــــعـليـم العـــــــالي و البـــــــــحث العـــــلـمـــي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جـــــــــــــامعة أبي بكـــر بـلـقـايد – تـلمســـــــــــــان –

Université Aboubakr Belkaïd – Tlemcen –

Faculté de TECHNOLOGIE

THESE

Présentée pour l'obtention du **grade** de **DOCTEUR EN SCIENCES**

En : Télécommunication

Spécialité : Télécommunication

Par : BOUYEDDOU Benamar

Sujet

Détection avancée des anomalies et applications aux systèmes de communication

Soutenue publiquement, le 03 Avril 2021, devant le jury composé de :

Mr. HADJILA Mourad	MCA	Univ. Tlemcen	Président
Mr. KADRI Benamar	Professeur	Univ. Tlemcen	Directeur de thèse
Mr. DAMOU Mehdi	MCA	Univ. Saida	Examinateur 1
Mr. BENMAMMAR Badr	Professeur	Univ. Tlemcen	Examinateur 2
Mr. DENNOUNI Nassim	MCA	Univ. Chlef	Examinateur 3
Mr. HARROU Fouzi	Chercheur Scientifique	KAUST –Arabie Saoudite	Invité

Année universitaire : 2020-2021

Remerciements

Tout d'abord, je tiens à remercier le bon Dieu le tout Puissant de m'avoir donné la force et le courage de mener à bien ce modeste travail

Je tiens à exprimer ma gratitude et mes vifs remerciements à mon directeur de thèse, Monsieur Benamar KADRI, Professeur à l'Université Abou Bekr Belkaïd de Tlemcen pour sa disponibilité, son encouragement, son expérience et ses conseils qui m'ont permis de mener à bien cette thèse.

Je suis très reconnaissante envers Monsieur Fouzi HARROU, mon co-directeur de thèse, pour sa disponibilité malgré la distance qui nous sépare, sa patience, sa générosité et les nombreuses discussions fructueuses qui ont animé ces quatre années de thèse.

Je suis très honoré par la présence de Monsieur Mourad HADJILA, qui a accepté de présider le jury de ma thèse, je suis également très honoré par la présence de Monsieur Mehdi DAMOU, Monsieur Badr BENMAMMAR et Monsieur Nassim DENNOUNI qui ont accepté d'être les rapporteurs de cette thèse. Qu'ils trouvent ici mes plus vifs remerciements pour l'effort qu'ils ont fait pour lire mon manuscrit et l'intérêt qu'ils ont porté à mon travail.

J'adresse mes sincères remerciements à Monsieur Fakhr Eddine HACHEMI, l'enseignant à l'université de Saida avec qui j'ai partagé le séjour, pour sa générosité, sa qualité humaine, et de faire partager son expérience, dont j'ai beaucoup appris.

J'exprime toute ma gratitude envers toutes les personnes qui m'ont aidés et soutenue de près ou de loin et ayant contribué au bon déroulement de cette thèse.

Je réserve une pensée particulière à ma famille, qui ont toujours été là pour moi.

Benamar Bouyeddou

Dédicace

A toute ma famille.

Résumé

La détection précoce des anomalies dans les systèmes et technologies de l'information et de la communication est une tâche cruciale. Diverses sources peuvent être à l'origine de différentes formes d'anomalies, comme les défaillances matérielles et logicielles, les changements dans la topologie des réseaux, la mise en place de nouvelles applications, ainsi que suite à des activités malveillantes et des tentatives de cyber-attaques. En effet, le concept a été extensivement traité par les acteurs du domaine numérique, et de nombreuses solutions ont été élaborées et basées sur différentes techniques telles que l'intelligence artificielle, l'apprentissage profond, la modélisation statistique et à base de connaissance.

Dans cette thèse, nous nous sommes intéressés à l'utilisation des cartes de contrôle statistiques et les mesures de similarité à la détection des cyber-attaques de dénie de service DOS et DDOS dans les réseaux IP.

Nous étudions les performances des cartes de contrôle Shewhart, CUSUM et EWMA dans la détection des différents types des attaques DOS et DDOS. Une étude comparative entre les trois cartes est ainsi présentée.

Nous proposons deux types de mécanismes de détection automatique et en temps réel d'anomalies à base de la distance CRPS. Le premier type concerne la détection paramétrique et combine les cartes Shewhart et EWMA avec la distance CRPS. Le deuxième type consiste en une méthodologie de détection non-paramétrique nommée CRPS-ES, basée sur un lissage exponentiel des mesures CRPS et une règle de décision via KDE. Ces mécanismes sont testés sous différents scénarios d'attaques DOS et DDOS. Les résultats obtenus montrent une amélioration significative des performances de détection.

Nous introduisons, ainsi, une procédure de détection ES-KLD basée sur la divergence KLD. ES-KLD implémente une adaptation temps réel de KLD, un lissage exponentiel des mesures KLD et un seuil de détection automatique, adaptable et qui ne nécessite pas de connaissance préalable de la distribution du trafic sous-jacente. Les performances de l'ES-KLD sont validées en présence de diverses formes des attaques DOS et DDOS, et des résultats prometteurs ont été obtenus.

Mots clés : détection d'anomalies, anomalies de trafic, attaques DOS et DDOS, cartes de contrôle, lissage exponentiel ES, distance CRPS, divergence KLD, KDE, bases de trafic DARPA99, MAWI, ICMPv6

Abstract

Early anomaly detection in information and communication systems and technologies is a crucial task. Various forms of anomalies can occur for different reasons, such as hardware and software failures, changes in the network's topology, the implementation of new applications, as well as due to malicious activities and cyber-attacks. Indeed, this concept has been extensively addressed by the digital actors, and many solutions have been developed and based on different techniques such as artificial intelligence, deep learning, statistical and knowledge-based modeling.

In this thesis, we focus on the use of statistical control charts and measures of similarity to the detection of denial of service DOS and DDOS cyber-attacks in IP networks.

We study the performance of three commonly used charts, namely Shewhart, CUSUM, and EWMA in detecting different types of DOS and DDOS attacks. A comparative study between the three charts is presented.

We propose two types of automatic and real-time anomaly detection mechanisms based on the CRPS distance. The first type relies on parametric detection and combines the Shewhart and EWMA charts with the CRPS distance. The second type consists of a non-parametric detection methodology called CRPS-ES, based on an exponential smoothing of CRPS measurements and a decision rule via KDE. These mechanisms are tested under different DOS and DDOS attack scenarios. The results show a significant improvement in detection performance.

We introduce an ES-KLD detection procedure based on KLD divergence. ES-KLD implements a real-time adaptation of KLD, an exponential smoothing of KLD measurements, and an automatic detection threshold that is flexible and does not require prior knowledge about the distribution underlying traffic. The performance of the ES-KLD is validated in the presence of various forms of DOS and DDOS attacks, and promising results have been obtained.

Keywords: anomaly detection, traffic anomalies, DOS and DDOS attacks, control charts, exponential smoothing ES, CRPS distance, KLD divergence, KDE, traffic datasets DARPA99, MAWI, ICMPv6

ملخص

إن الكشف المبكر عن حالات الخلل في نظم وتكنولوجيات المعلومات والاتصالات مهمة بالغة الأهمية. عمليا يمكن أن يكون هناك مصادرمختلفة و التي تنتج بدورها أشكال مختلفة من الخلل ، مثل فشل الأجهزة والبرامج، التغيرات في طوبولوجيا الشبكات، وتنفيذ تطبيقات جديدة، فضلا عن الأنشطة الخبيثة ومحاولات الهجمات السيبرانية. والواقع أن الجهات الفاعلة في المجال الرقمي قد تناولت هذاالموضوع على نطاق واسع، أين تم تطوير العديد من الحلول والتي استندت إلى تقنيات مختلفة مثل الذكاء الاصطناعي والتعلم العميق والنمذجة الإحصائية والمعارف .

في هذه الأطروحة، ندرس استخدام خرائط التحكم الإحصائية ومقاييس التشابه للكشف عن هجمات حجب الخدمة DOS وDDOS في شبكات IP.

ندرس أداء خرائط التحكم Shewhart و CUSUM و EWMA في الكشف عن أنواع مختلفة من هجمات DOS و DDOS ونعرض دراسة مقارنة بين الخرائط الثلاث.

نطور نوعين من آليات الكشف التلقائية والزمن الحقيقي عن الخلل القائم على اساس CRPS.
النوع الأول يتعلق بالكشف البارامتري ويجمع بين بطاقات Shewhart و EWMA مع مسافة CRPS.
ويتألف النوع الثاني من منهجية كشف غير بارامتري تسمى CRPS-ES ، استناداً إلى سلاسة أسية لقياسات الـ CRPS وقاعدة قرار عبر KDE .يتم اختبار هذه الآليات ضمن سيناريوهات هجوم DOS و DDOSمختلفة. تظهر النتائج المتحصل عليها تحسنا كبيرا في أداء الكشف.

نقدم طريقة الكشف ES-KLD على أساس الاختلاف KLD .ES-KLD ينفذ التكيف في الوقت الحقيقي منKLD، و إلى سلاسة أسية لقياسات KLD وعتبة كشف تلقائية قابلة للتكيف ولا ةتطلب معرفة مسبقة لدالة توزيع احتمال traffic . تم التحقق من أداء ES-KLD في وجود أشكال مختلفة من هجمات DOS وDDOS، وقد تم الحصول على نتائج واعدة .

الكلمات المفتاحية: الكشف عن الخلل، خلل traffic، هجمات حجب الخدمة و حجب الخدمة الموزعة ، خرائط التحكم الإحصائية ، سلاسة أسية ، مسافة CRPS، الاختلاف KLD ، KDE ,قواعد بيانات traffic, ICMPv6, MAWI, DARPA99.

Tables des Matières

Chapitre 2

Mitigation des cyber-attaques par les techniques de détection d'anomalies

Chapitre 3

Détection d'anomalies via les cartes de contrôle : application à la détection des cybers-attaques DOS et DDOS dans les réseaux IP

Chapitre 4

Techniques de détection d'anomalies à base de la distance CRPS

Chapitre 5

Méthode non-paramétrique de détection d'anomalies à base de la divergence Kullback-Leibler

Liste des figures

Chapitre 1

Chapitre 2

Chapitre 3

Chapitre 4

Chapitre 5

Liste des tableaux

Chapitre 3

Chapitre 4

Chapitre 5

Liste des abréviations

A

ACK: ACKnowledgement

ACL: Access Control List

AFRL: Air Force Research Laboratory

AIS: Artificial Immune System

A-NIDS: Anomaly-based Network Intrusion Detection Systems

ARP: Address Resolution Protocol

AS: Autonomous System

ASFI : Attaque SYN à Faible Intensité

ASIE : Attaque SYN d'Intensité Elevée

B

BN: Bayesian Network

C

CIA: Confidentiality/Integrity/Availability

CDF : Cumulative Density Function

CL: Central Line

CPN: Colored Petri Nets

CPS: Cyber-Physical Systems

CPU: Central Processing Unit

CUSUM: CUmulative SUM

CRPS: Continuous Ranked Probability Score

D

DAD: Duplication Address Detection

DARPA99: Defense Advanced Research Projects Agency 99

DE: Decision Engine

DIDS: Dependable network IDS

DOS: Denial Of Service

DDOS: Distributed DOS

DNS: Domaine Name Server

E

ES: Exponential Smoothing

EWMA: Exponentially Weighted Moving Average

F

FN: False Negative

FNR: True Negative Rate

FP: False Positive

FPR: False Positive Rate

FTP: File Transfer Protocol

G

GA: Genetic Algorithm

GMM: Gaussian Mixture Model

Go: Giga octet

H

HD: Hellinger Distance

HIDE: Hierarchical Intrusion DEtection

HTTP: Hypertext Transfer Protocol HTTP

I

ICMP: Internet Control Message Protocol

ICMPv4: Internet Control Message Protocol version 4

ICMPv6: Internet Control Message Protocol version 6

IDA: Intrusion detection Agent

IDEVAL: Intrusion DEtection EVALuation

IDIOT: Intrusion Detection In Our Time

IDS: Intrusion Detection System

IGMP: Internet Group Management Protocol

IIS: Internet Information Server

IoT: Internet of Things

IP: Internet Protocol

IPS: Intrusion Prevention System

IPv4: IP version 4

IPv6: IP version 6

K

KDE: Kernel Density Estimator

KDD99: Knowledge Discovery and Data mining 99

KLD: Kullback-leibler Divergence

KNN: K-Nearest Neighbour

L

LAND: Local Area Network Denial

LCL: Lower Control Limit

LR-DOS: Low Rate-Denial Of Service

LR-DDOS: Low Rate-Distributed Denial Of Service

M

MAWI: Measurement and Analysis on the WIDE Internet

MIC: Maximum Information Coefficient

MIT: Massachusetts Institute of Technology

Mo: Méga octet

MSPCA : Multi-Scale Principal Component Analysis

N

NA: Neighbor Advertisement

NMAP: Network MAPper

NN: artificial Neural Network

NS: Neighbor Solicitation

NTP: Network Time Protocol

P

PAIDS: Proximity-Assisted IDS

PCA: Principal Composant Analysis

PDF: Probability Density Function

PHP: Hypertext Preprocessor

R

RAM: Random Access Memory

S

SIP: Session Initiation Protocol

SIP INVIT: Session Initiation Protocol INVITation

SMPT: Simple Mail Transfer Protocol

SMS: Short Message Service

SNMP : Simple Network Management Protocol

SPC: Statistical Process Control

SQL: Structured Query Language

SSH: Secure Shell

SunOS: Sun Operating System

SVM: Support Vector Machine

SYN: SYNchronization

T

Tbps: TeraBits Per Second

TCP: Transmission Control Protocol

Telnet: Terminal over a network

TIC : Technologies de l'information et des systèmes de Communication

TN: True Negative

TNR: True Negative Rate

TP: True Positive

TPR: True Positive Rate

U

UCL: Upper Control Limit

UDP: User Datagram Protocol

W

WSN: Wireless Sensor Network

X

XSS: Cross-site scripting

Introduction générale

Introduction générale

Au cours des deux dernières décennies, l'émergence des technologies de l'information et des systèmes de communication (TIC) a considérablement changé la façon dont les informations sont accessibles et communiquées, notamment via les réseaux IP (Internet Protocol). Cette évolution des TIC a également entraîné une dépendance de la société à l'égard des systèmes de stockage, de traitement et de partage de l'information. La gamme de services pris en charge ne cesse de se multiplier, en incluant une pléthore d'applications des objets connectés (l'internet des objets) et le contrôle et la surveillance des systèmes sensibles et à infrastructure critique, tels que l'électricité, l'eau, le gaz, …etc. [1].

Par conséquence, les éventuels dysfonctionnements (anomalies) dans ces systèmes peuvent causer de légers inconvénients, une perte de productivité, un déficit économique, voire de dommages graves, irréparables et un état de mal-être public. Si certains résultent des erreurs de configuration, défaillance de matériels, plantage de serveurs, pagination réseau, flux de diffusion et les situations de congestion; ce sont les anomalies liées aux problèmes de sécurité qui préoccupent de plus en plus les acteurs de l'aire du numérique [2]. En effet, ces anomalies découlent généralement des activités intentionnellement malicieuses, initiées par les cybers attaquants pour démonter le triangle CIA (Confidentiality, Integrity, Availability) d'un système TIC victime. Dans ce sens, les attaques de dénie de service DOS (Denial Of Service) et leur forme distribuée DDOS (Distributed DOS) s'inscrivent parmi les menaces majeures pour les différentes architectures réseaux. Elles implémentent de mécanismes extrêmement puissants, et tentent à interrompre temporairement, voire définitivement les services offerts par un système TIC victime. C'est pourquoi il est autant important de mettre en place des solutions adéquates pour sécuriser les actifs et les protéger contre les effets dévastateurs de ces attaques. Tout comme les autres outils de sécurité (pare-feu, antivirus, systèmes de contrôle d'accès), les mécanismes de détection des attaques DOS et DDOS essayent de renforcer le niveau de sécurité et maintenir un état de stabilité approprié.

■ Motivations :

Au fil des ans, les techniques de détection des attaques DOS et DDOS ont beaucoup évolué pour y faire face. En particulier, les techniques de détection à base d'anomalies

gagnent plus d'importance et sont reconnues comme étant plus efficaces, notamment dans la détection de nouvelles attaques. Toutefois, le problème est loin d'être résolu et de nombreuses opportunités offrent d'avantages pour plus de progrès, avant qu'elles soient déployées à grande échelle. En tenant compte de leurs performances prometteuses, les techniques statistiques constituent actuellement l'un des principaux axes de recherche dans ce domaine. Elles sont principalement caractérisées par leurs capacités inhérentes à fournir de taux de détection élevés. En outre, l'essor des capacités de calcul, leur permet de notifier les résultats de détection en temps réel; facteur fondamental pour détecter les attaques à un stade précoce. Par ailleurs, ils restent de problèmes clés à résoudre avant qu'elles peuvent être déployées à grande échelle. Ces limitations incluent :

- Le problème des fausses alarmes FP (False Positive) : comme toutes les techniques à base d'anomalies, le taux des FP est souvent plus élevé que dans les techniques à base de signatures.
- La limite de détection (Detection Threshold) : l'état normal des paramètres et des métriques doit être judicieusement choisi afin d'établir le meilleur compromis entre les faux positifs et les faux négatifs (FN : False Negative).
- La distribution du trafic : la plupart des techniques reposent sur l'hypothèse d'un processus normal, ce qui n'est pas toujours le cas.

■ Objectifs:

Dans cette thèse, nous donnerons un ensemble de solutions pour remédier aux limitations citées ci-dessus. Nous utiliserons les techniques de détection d'anomalies statistiques ainsi proposées pour la détection des différents types d'attaques DOS et DDOS.

Ces solutions concernent principalement :

- La mise en place des mécanismes qui optimisent le taux de détection et réduisent le taux des fausses alarmes.
- L'établissement de seuils de détections automatiques et évolutives en fonction de l'état de trafic et du réseau.
- Proposition de techniques adaptées aux trafics non Gaussiens via une évaluation non-paramétrique des distributions de trafic.

■ Contributions :

Nos travaux de recherches menés dans cette thèse pointent sur l'utilisation des cartes de contrôle et les mesures de similarité (distances et divergences) pour la détection des attaques et des intrusions dans les réseaux IPv4 (IP version 4) et IPv6 (IP version 6). Des méthodes statistiques de détection d'anomalies sont ainsi proposées. Nos principales contributions sont :

- Techniques de détection statistiques à base d'anomalies des attaques DOS et DDOS, en utilisant les cartes de contrôle Shewhart, EWMA (Exponentially Weighted Moving Average) et CUSUM (CUmulative SUM). Ces cartes déterminent automatiquement les limites de détection et leurs caractéristiques statistiques annoncent la présence ou l'absence des attaques dans le trafic contrôlé.
- Proposition, introduction et première utilisation de la distance CRPS (Continuous Ranked Probability Score) dans le domaine de la détection. CRPS est une méthode statistique très connues, et est largement utilisée dans les prévisions (forecasting) météorologiques et économiques. Nous allons l'utiliser pour quantifier la dissimilarité entre les mesures des trafics capturés et la distribution du trafic normal. Les valeurs importantes de cette métrique reflètent la présence des attaques DOS et DDOS. Ici, les limites de détection sont fixées par un lissage exponentiel (ES : Exponential Smoothing) ou par application des cartes de contrôle Shewhart et EWMA.
- Utilisation de la divergence KLD (Kullback-leibler Divergence) pour la détection des attaques DOS et DDOS. On propose une nouvelle implémentation dotée de nouvelles fonctionnalités, adaptée aux applications temps réel, avec un lissage exponentiel ES et de règle de décision non-paramétrique.
- Développement de nouvelles cartes de contrôle qui combinent entre les cartes de contrôle conventionnelles, la distance CRPS. Les cartes proposées sont : CRPS-EWMA et CRPS-Shewhart. Précisément, pour révéler les susceptibles attaques, EWMA et Shewhart seront appliquées aux mesures CRPS et non pas aux mesures des trafics capturés.
- Il n'y a aucune hypothèse quant à la normalité des données et une estimation plus réaliste de la distribution du trafic est proposée. Nous avons introduit une estimation non-paramétrique des distributions de probabilités via KDE (Kernel Density Estimator).

- Développement d'une méthodologie de prétraitement (pre-processing) des traces de trafic. Par exemple, pour DARPA99, nous avons passé d'une trace de 8Go à une nouvelle base de quelques Mo.

Organisation de la thèse

La suite de ce document s'organise comme suit :

Dans le premier chapitre, nous introduisons les concepts des attaques de dénie de service DOS et DDOS. Nous commençons d'abord par une présentation générale de ces attaques et leurs éventuelles cibles dans un réseau IP. Nous donnons leur classification en décrivant les différentes techniques utilisées pour générer ces types d'attaques. Nous décrivons par la suite les attaques DOS et DDOS les plus pertinentes. On finira le chapitre par un résumé sur les solutions et recherches menées dans la détection des attaques DOS et DDOS.

Dans le deuxième chapitre, nous présentons un état de l'art sur l'utilisation des techniques de détection d'anomalies dans la détection des intrusions et des attaques DOS et DDOS dans les réseaux IP. Après la définition de la notion de la détection d'anomalie, nous exposons les différents types des anomalies réseaux. Nous décrivons ensuite les modes de détection ainsi que les mesures utilisées pour les évaluer. Nous détaillons aussi la structure générale que doit avoir un système de détection à base d'anomalies. Enfin, nous introduisons une classification générale de ces techniques, en présentant les travaux effectués dans ce sujet.

Dans le troisième chapitre, nous investiguons l'utilité des cartes de contrôle dans la détection des attaques DOS et DDOS. Nous commençons par introduire le principe de base de ces cartes, en donnant une description générale des cartes les plus utilisées, à savoir Shewhart, CUSUM et EWMA. Ensuite, nous détaillons la procédure adoptée pour utiliser ces cartes dans la détection des attaques DOS et DDOS, suivie par une évaluation de leurs performances sous différents types d'attaques, et une étude comparative entres les trois cartes est ainsi menée.

Dans le quatrième chapitre, nous proposons l'utilisation de la distance CRPS dans le domaine de la détection d'anomalies et la détection des cyber-attaques de type DOS et DDOS. Nous commençons par une présentation générale de la distance CRPS, en présentant nos motivations et les objectifs voués par cette proposition. Ensuite, nous décrivons le principe d'utilisation de la CRPS pour la détection d'anomalies et des attaques DOS et DDOS. Ici, nous proposons les cartes de contrôle à base CRPS, et qui sont CRPS-Shewhart,

CRPS-EWMA et CRPS-ES. Pour les valider, leurs performances sont évaluées avec différentes traces de trafic réseau pour IPv4 et IPv6.

Dans le cinquième chapitre, nous présentons une technique de détection des attaques DOS et DDOS à base la divergence KLD. Après l'introduction de la KLD, nous détaillons la méthodologie mise en place pour l'utiliser dans la détection d'anomalies et des attaques DOS et DDOS et la nouvelle implémentation ES-KLD ainsi proposée. Ensuite, le mécanisme ES-KLD sera validé sous différents scénarios des attaques DOS et DDOS fournis par différentes bases de trafic réseau pour les environnements IPv4 et IPv6.

Enfin, nous concluons la thèse par une synthèse générale des contributions et des résultats obtenus, ainsi que les potentielles perspectives des travaux effectués.

Chapitre1

Les cyber-attaques

DOS et DDOS dans les réseaux IP

Dans ce chapitre

- Les cyber-attaques de dénie de service DOS et DDOS
- Les cibles des attaques DOS et DDOS
- Classification des attaques DOS et DDOS
- Exemples d'attaques DOS et DDOS
- Détection des attaques DOS et DDOS

Chapitre 1

Les cyber-attaques DOS et DDOS dans les réseaux IP

Depuis son apparition, l'Internet a révolutionné notre quotidien. Il serait insensé de recenser les innombrables services qui sont devenus possibles grâce à l'Internet et les technologies associées (Cloud, Internet of Things IoT,...). Des milliers d'organisations telles que les passerelles de paiement, les moteurs de recherche (Google, Yahoo), les banques, les instituts éducatifs, les serveurs commerciaux (Flipkart, eBay, Amazon), les sites web des réseaux sociaux (Twitter, Facebook), les bourses, les prévisions météorologiques,… ont déployé leur serveurs pour fournir des services et des applications diverses aux grand public [3].

En contrepartie, avec l'omniprésence de l'Internet et le nombre hyper important de machines connectées, les cyber-attaques ont aussi connues une forte progression. Parmi ces attaques, les attaques de dénie de service DOS et DDOS tendent de devenir de plus en plus les types les plus répandus. En fait, elles provoquent des coûts financiers importants liés à la restauration, l'interruption opérationnelle, pertes de revenues, perte de productivité, atteinte de réputation et d'autres formes du cyber-chaos (dommages économiques pour les organisations et les particuliers). Auprès la gravité des leurs conséquences, ces attaques ont été largement étudiées par la communauté scientifique, et de nombreux mécanismes de défense ont été élaborés pour les neutraliser.

Ce chapitre constituera une partie introductive de la problématique des attaques de dénie de service DOS et DDOS dans les réseaux IP. On commencera par leur définition et principe de fonctionnement. On citera ensuite leurs principales cibles. Nous introduirons leur classification générale, suivie par une description des quelques exemples, et on terminera par un aperçu général sur les mécanismes de défense contres ces attaques.

1.1. Les attaques de dénie de service DOS et DDOS

Ces dernières années, diverses organismes de cyberdéfense affirment que les incidents liés aux attaques de dénie de service ont connu une croissance effrayante. Comme le montre la figure 1.1, des dizaines (voire des centaines) d'attaques sont enregistrées chaque jour. Selon Arbor Networks (2018), le fournisseur de services américain Github a survécu à une attaque qui a atteint un débit record sans précédent de 1,7 Tbps [4].

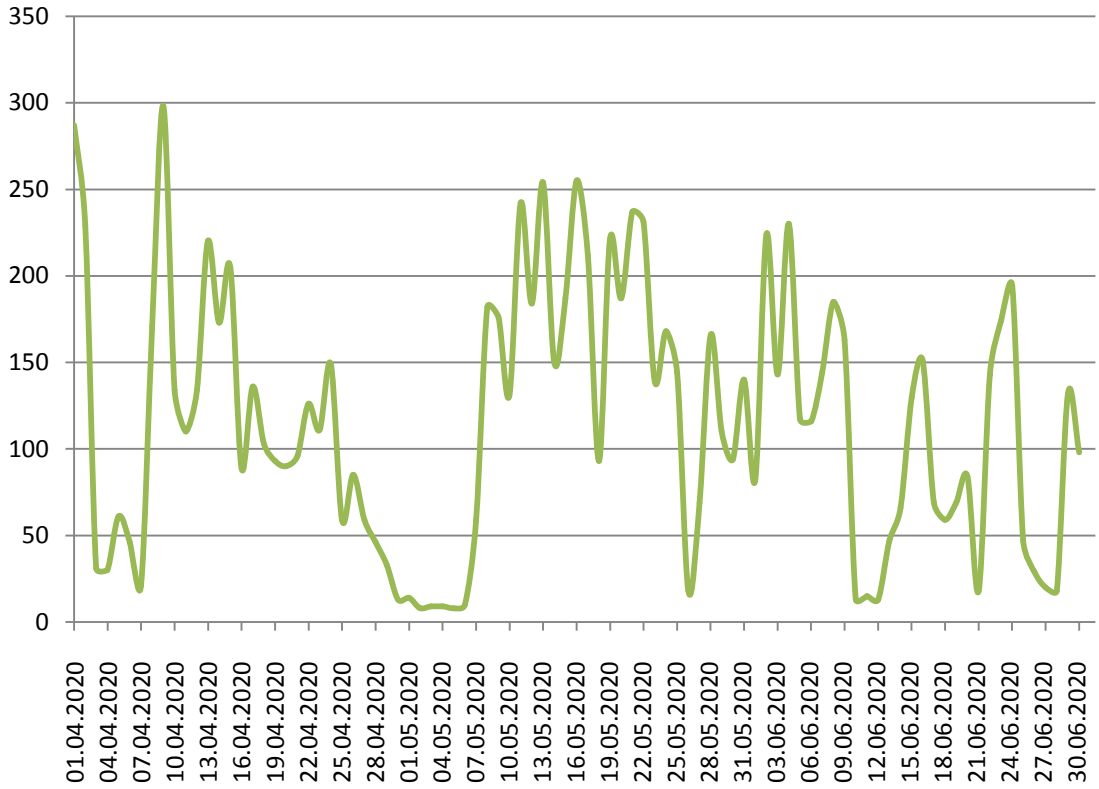

Figure 1.1 : Evolution des attaques DDOS entre Avril et Juin 2020 [5]

Ces attaques sont destinées à perturber, voire rendre totalement inaccessible un service, un serveur ou une infrastructure réseau. Ils peuvent se manifester sous plusieurs formes, mais généralement elles tentent à épuiser les ressources de leurs victimes, à savoir la bande passante, capacité de traitement (CPU : Central Processing Unit) et la capacité mémoire (RAM : Random Access Memory) [6]. Une attaque DOS (figure 1.2) se fait générée depuis une seule source, alors qu'une DDOS (figure 1.3) implique généralement un très grand nombre d'hôtes. Ces hôtes peuvent être des amplificateurs, des réflecteurs ou des bots (zombies) synchronisant leurs attaques suite à une commande de contrôle issue de l'attaquant.

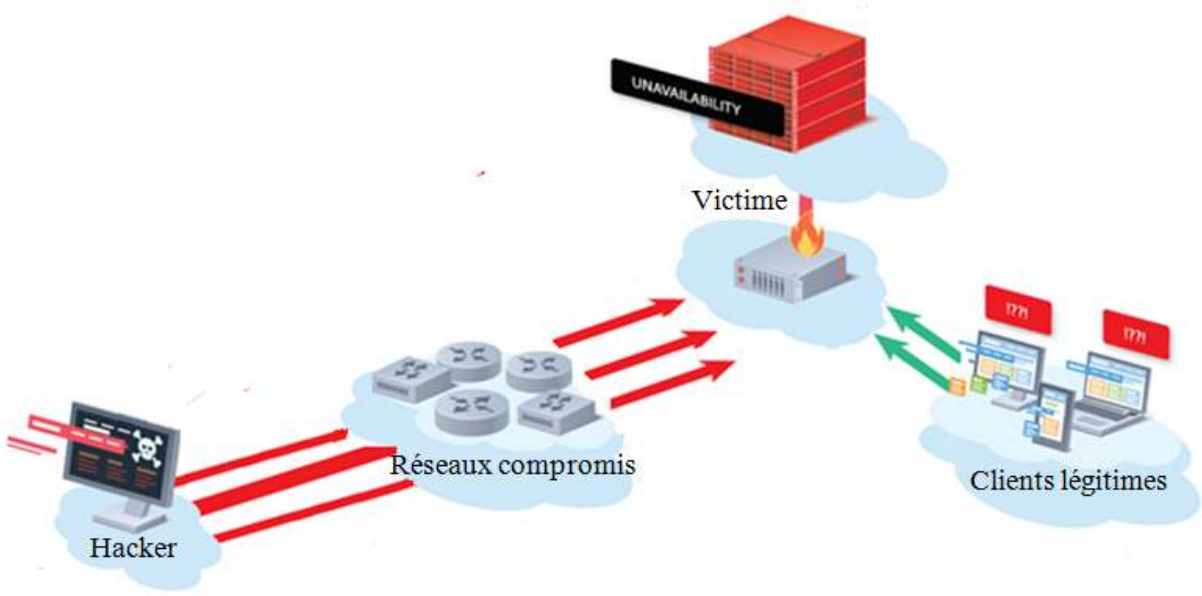

Figure 1.2 : Attaques de dénie de service DOS

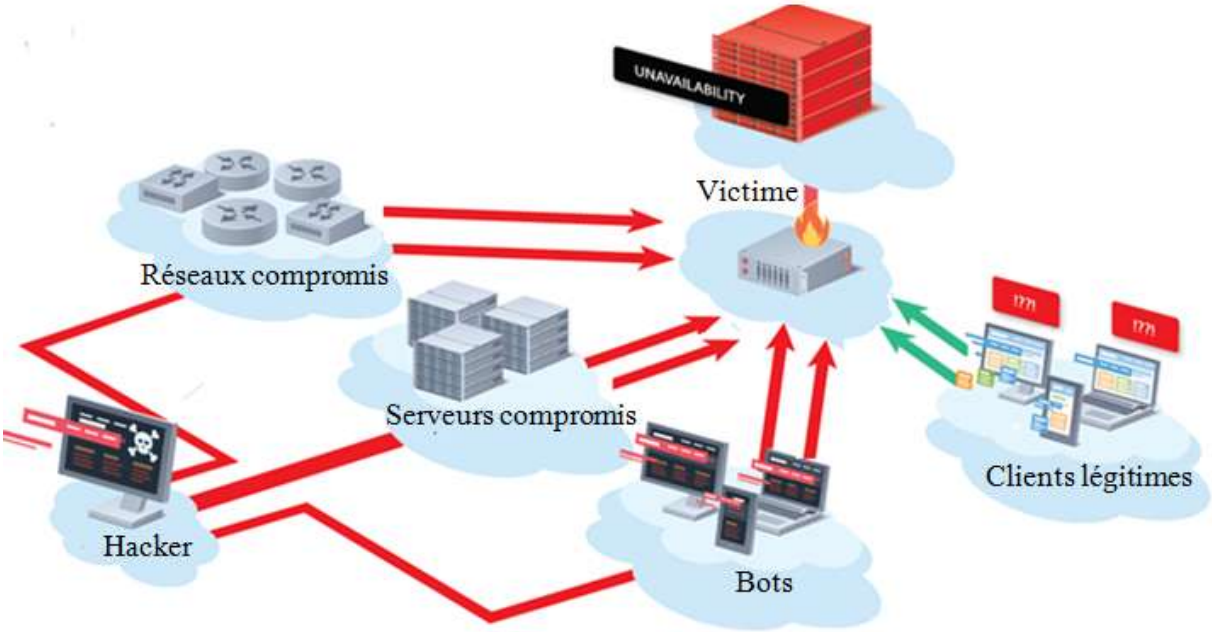

Figure 1.3 : Attaques de dénie de service distribuées DDOS

1.2. Les cibles des attaques DOS et DDOS

Tous les actifs d'une infrastructure réseau connectée à l'Internet sont vulnérables aux attaques DOS et DDOS. Le RFC 4732 [7] cite les principales cibles, notamment :

- Les systèmes d'extrémité sont les victimes les plus courants, tels que les serveurs web, les pare-feu, les IDS (Intrusion Detection System), les DNS (Domaine Name Server) ou tout autre hôte normal. Pour épuiser les ressources de ces systèmes, les attaques

DOS et DDOS s'appuient généralement sur les vulnérabilités des logiciels, des applications et des systèmes d'exploitation.

- Les routeurs: les attaques DOS et DDOS qui visent les routeurs peuvent s'avérer très dangereuses. Car, empêcher un routeur d'offrir les services pour lesquelles il est destiné, il le met dans un état critique et de nombreux utilisateurs comme d'autres routeurs pourraient devenir isolés. Par conséquence, de nombreux services ne seront plus accessibles. En outre, suite à une attaque DOS ou DDOS, le système autonome AS (Autonomous System) qui appartient au domaine de routage peut tomber en panne, si le routeur attaqué était une passerelle vers un sous-réseau, ou s'il s'agit d'un routeur connecté à d'autres routeurs dans une certaine zone ou dans d'autres zones. Un routeur peut être attaqué comme tout système d'extrémité, ou une par attaque spécifique via les protocoles de routage. Une attaque DOS/DDOS peut cibler le processeur de contrôle du routeur et lui rendre complètement ingérable. Cela peut empêcher aussi la prise de mesures susceptibles d'atténuer l'attaque et empêcher le diagnostic de la cause du problème.
- Liens : envoyer suffisamment de trafic non contrôlé de sorte qu'une liaison devient énormément encombrée, et le trafic légitime subira un taux de perte de paquets trop élevé. Pour l'accès sans fil, des conflits et collisions fréquents peuvent avoir lieu à cause d'une attaque de dénie de service. Il faut noter ici que tout type de dommage physique peut être considéré comme une attaque, par exemple, couper un câble d'alimentation ou détruire une liaison d'accès arrêtera définitivement un système et les services qui s'exécutent dessus.
- Connexions actives : au lieu d'attaquer le système d'extrémité lui-même, il est également possible pour un attaquant de perturber les communications en cours. En usurpant des paquets dans la connexion, il peut la réinitialiser, la désynchroniser, la suspendre et effectuer d'autres formes d'intrusions. Ainsi, un utilisateur malveillant pourrait être en mesure de réduire considérablement le débit d'une connexion en cours en envoyant des messages falsifiés de type ICMP (Internet Control Message Protocol) extinction de la source.
- Tout service permettant l'envoie ou la réception des données est vulnérable aux attaques DOS/DDOS, par exemple DOS sur les réseaux de capteurs sans fil (WSN : Wireless Sensor Network), et les systèmes IoT.

1.3. Classification des attaques DOS et DDOS

La littérature des attaques DOS et DDOS compte déjà plusieurs classifications [8] [9] [10]. Nous considérons ici une classification basée sur leurs impact sur le réseau ou les ressources des victimes qu'elles ciblent [10]. En général, les hackers exploitent les failles de telle sorte que les ressources affectées aux systèmes victimes (ex : RAM et CPU) soient débordées et ne peuvent pas poursuivre leurs opérations normales. Par conséquence, la victime finit par rejeter les requêtes des clients. Comme ils peuvent épuiser les bandes passantes réseau. Dans ce cas, ils produisent un flux important de trafic malveillant pour submerger la bande passante du réseau, ce qui affectera non seulement la victime, mais aussi tout actif branché au chemin de cette attaque y inclus d'autres systèmes connectés à ce réseau. Ainsi, en tenant compte de ces types d'impact, les attaques DOS et DDOS peuvent être groupées en deux grandes classes : les attaques d'épuisement de la bande passante et les attaques d'épuisement des ressources. Toutefois, en pratique, une attaque peut avoir les deux impacts, imposant les effets les plus importants. En parle dans ce cas des attaques d'infrastructure [10]. Ces classes ainsi que les sous-classes qu'elles dérivent sont récapitulées dans la figure 1.4. Ci-après, une brève description des principes de ces attaques est ainsi présentée [10] [11] [12].

1.3.1. Attaques d'épuisement de ressources

Le but des attaques d'épuisement de ressources est de déborder ou de planter toutes les principales ressources du système victime telles que la mémoire, les sockets et le CPU. Deux manières différentes sont possibles pour mettre en marche ce type d'attaques. Dans un premier temps, l'attaquant exploite les failles de certains protocoles des couches réseau, transport et application. Dans le second cas, des paquets malformés sont utilisés pour effectuer les attaques.

- **Attaques par exploitation de protocoles**

Il existe plusieurs variantes d'attaques qui exploitent les faiblesses des différents protocoles réseau. Cela force la victime à utiliser entièrement tout son CPU et sa RAM pour effectuer certaines opérations, souvent très lourdes. Par exemple, les attaques de ce groupe exploitent les protocoles de la couche transport tels que TCP (Transmission Control Protocol)

et UDP (User Datagram Protocol), certains protocoles de la couche application tels que HTTP (Hypertext Transfer Protocol HTTP) et SIP (Session Initiation Protocol).

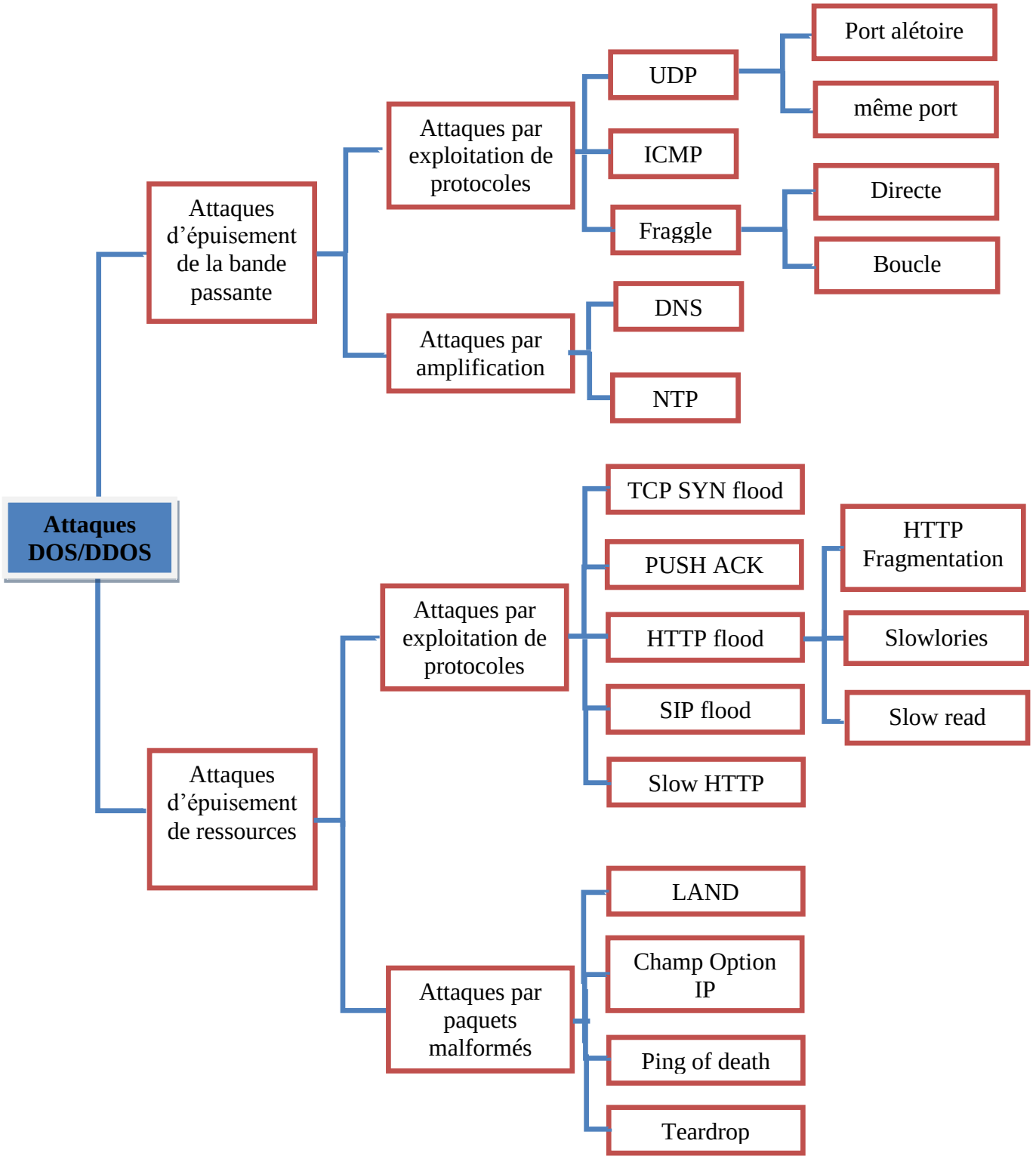

Figure 1.4: Différentes classes des attaques DOS/DDOS

- **Attaques par paquets malformés**

L'idée principale derrière une attaque par paquet malformé est d'attaquer une victime en utilisant un paquet déformé qui peut confondre la victime et par conséquent provoquer un plantage du système. Par exemple, l'attaque LAND (Local Area Network Denial) est basée sur des paquets dont l'adresse IP source et destination sont identiques, et l'attaque Teardrop utilise des fragments avec des déplacements erronés [10].

1.3.2. Attaques d'épuisement de la bande passante

L'attaque d'épuisement de la bande passante est une autre variante importante des attaques DOS/DDOS. Ici, le but de l'attaquant est de consommer toute la bande passante réseau du système victime. En effet, la victime décline les requêtes des utilisateurs légitimes temporairement voire définitivement, jusqu'à ce que l'attaque soit atténuée. Les attaques apparentant à cette catégorie peuvent être également déclenchées par l'exploitation des protocoles, comme elles peuvent être réalisées par amplification où un réflecteur ou un amplificateur est impliqué pour augmenter l'intensité d'attaque ainsi que les dommages au réseau victime [10] [11].

- **Attaques par exploitation de protocoles**

Suivent le même principe des attaques d'épuisement de ressources, elles peuvent exploiter un protocole de la couche transport tel que le protocole UDP ou un protocole de la couche réseau tel que l'ICMP [11].

- **Attaques par amplification**

L'idée principale derrière ce type d'attaques est de générer une réponse volumineuse pour une très petite requête et la diriger vers la victime qui finit par consommer toute sa bande passante. Deux attaques d'amplification très similaires et très répandues sont l'attaque par amplification DNS et l'attaque par amplification NTP (Network Time Protocol) [10] [12].

1.4. Description de quelques exemples d'attaques DOS/DDOS

Les travaux de recherche menés dans cette thèse se concentrent sur les attaques DOS et DDOS par inondation. Nous allons décrire dans ce paragraphe quelques exemples de ces

attaques. Ces attaques sont largement utilisées à l'heure actuelle, comme l'illustre la figure 1.5. Ainsi, nos contributions dans les chapitres 3, 4 et 5 seront validées à travers des traces de trafic IPv4 et IPv6 qui incluent ces attaques, à savoir les attaques TCP SYN flood, UDP flood, SMURF et les attaques basées sur les messages ICMPv6 sollicitation de voisins, annonce de voisin et annonce de routeur .

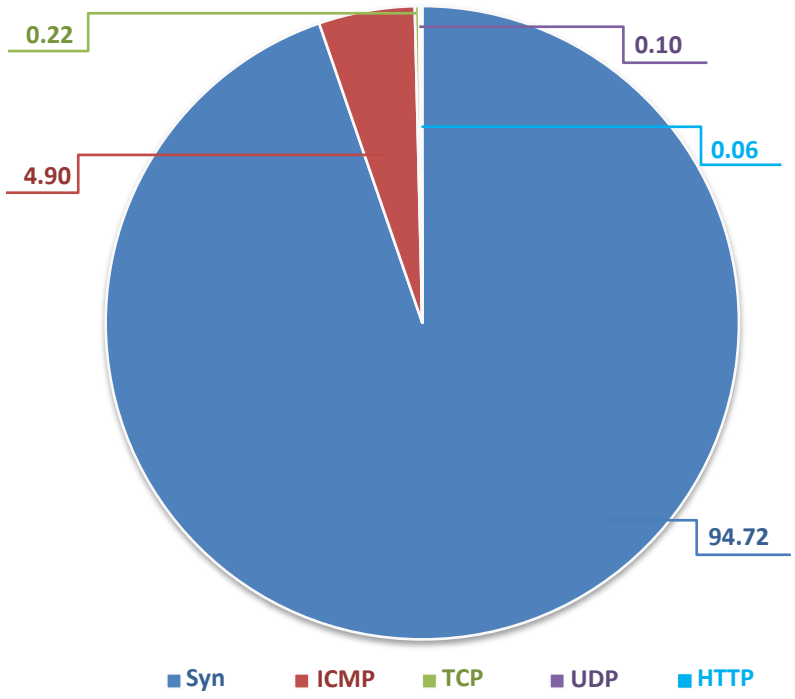

Figure 1.5: Distribution des attaques DOS/DDOS par type, Trimestre 2, 2020 [5]

1.4.1. Attaque TCP SYN flood

L'attaque TCP SYN flood reste toujours l'attaque la plus répandue; elle était utilisée dans 94.72% des attaques enregistrées dans le deuxième trimestre de 2020 (figure 1.5) [5]. Elle est couramment utilisée pour effectuer des attaques DOS ou DDOS contre tous les services basés sur le protocole TCP tels que les serveurs web, les serveurs FTP (File Transfert Protocol) ou les serveurs de messagerie, et qui exploite la faiblesse du mécanisme d'établissement à trois étapes d'une connexion TCP en conservant certaines ressources de backlog pour maintenir les connexions semi-ouvertes.

Lors d'une connexion TCP normale (figure 1.6), le client demande, une nouvelle connexion, en envoyant un segment SYN (SYNchronization) au serveur. Pour accuser la

réception de cette demande, le serveur répond avec un segment SYN/ACK (SYNchronization/ACKnowledgement) et ajoute cette demande à la file d'attente de backlog, intégrée dans sa mémoire système pour maintenir toutes les connexions semi-ouvertes. Ensuite, le client retourne au serveur un segment ACK de confirmation. Enfin, la connexion est établit et le serveur enlève la demande de la file d'attente du backlog. Toute demande de connexion restera dans la file d'attente du backlog jusqu'à ce que le serveur reçoive l'ACK du client, ce qui signifie que si, pour une raison quelconque, le serveur ne reçoit pas le segment ACK, la connexion reste ouverte (en occupant certaines ressources dans la file d'attente du backlog) pendant une période allant jusqu'à l'expiration de la connexion TCP (généralement environ 75 s) [13]. C'est exactement le point d'entrée exploité par les attaquants pour déclencher une inondation par segments SYN ; c'est l'attaque TCP SYN flood.

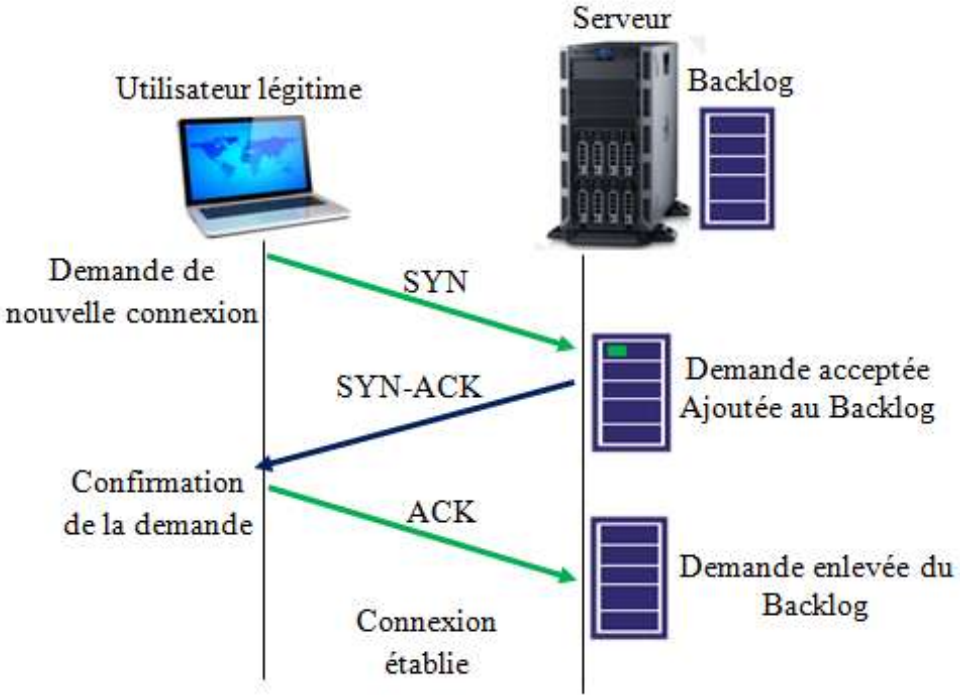

Figure 1.6: Procédure d'établissement d'une connexion TCP

Lors d'une attaque TCP SYN flood, l'attaquant crée un grand nombre de connexions semi-ouvertes, suffisamment pour atteindre la limite de file d'attente de backlog. Ainsi, le serveur victime ne peut pas accepter davantage de demandes de connexion, même celles des clients légitimes, et les services fournis par ce serveur sont donc suspendus. Dans la pratique, deux variantes de l'attaque TCP SYN flood peuvent être implémentées [14] [15]:

i) L'attaquant génère plusieurs segments SYN sans envoyer le segment ACK de confirmation. En fait, il ignore simplement le SYN/ACK reçu (figure 1.7).

ii) L'attaquant utilise une adresse IP source usurpée pour se connecter au serveur victime. Ce dernier enverra le SYN/ACK à l'adresse source usurpée. Etant donné que la source usurpée est inaccessible ou n'a pas envoyé la demande de connexion, le serveur ne recevra jamais l'ACK. Il y aura d'autant connexions semi-ouvertes que le nombre de SYN reçus.

Dans sa version DDOS (figure 1.8), l'attaquant compromet de nombreuses machines bots, puis leur ordonne d'attaquer, en même temps, le serveur victime. Chaque bot peut lancer d'ailleurs une attaque DOS TCP SYN au moyen des méthodes i) et ii), citées ci-dessus [15].

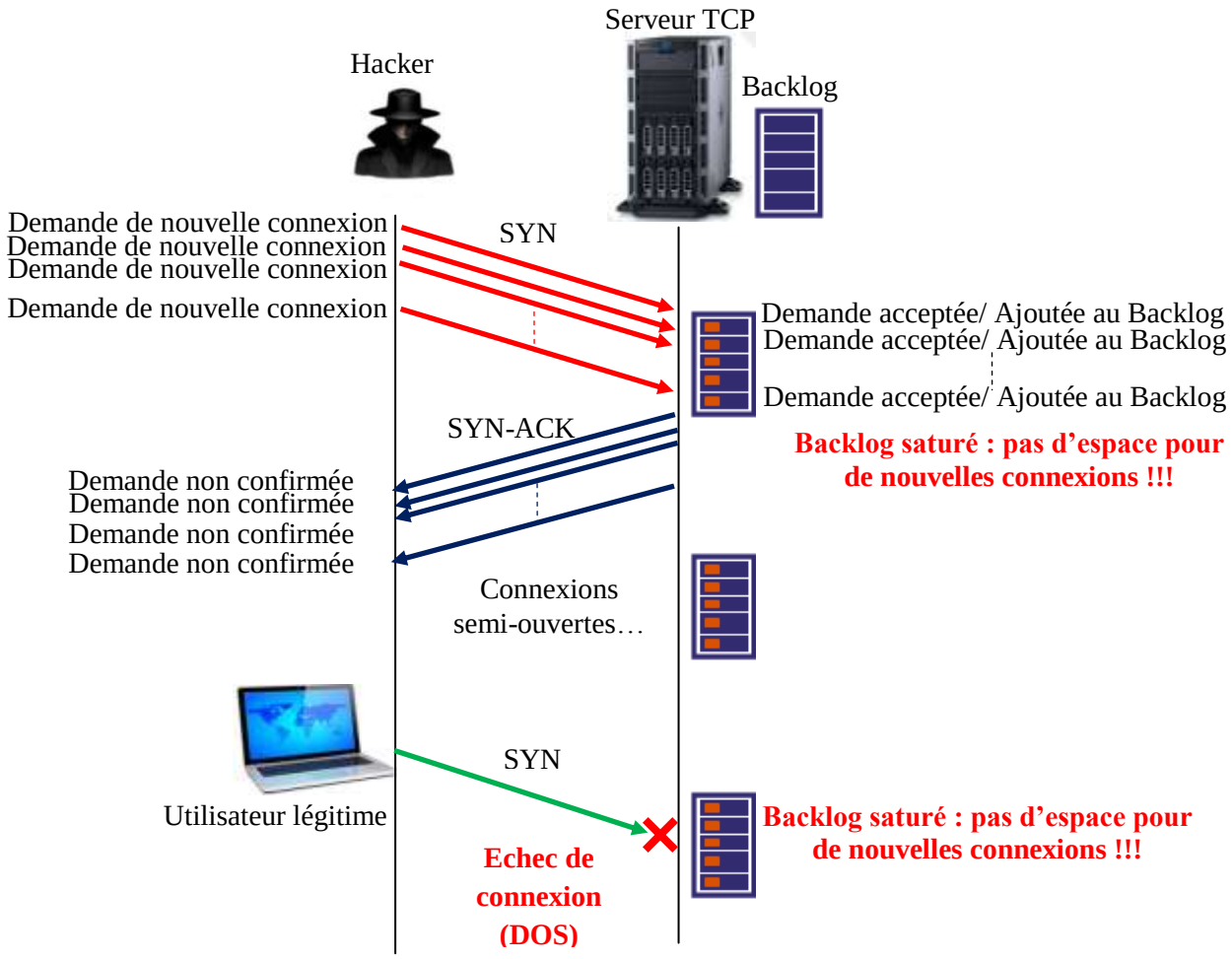

Figure 1.7: Attaque DOS TCP SYN flood

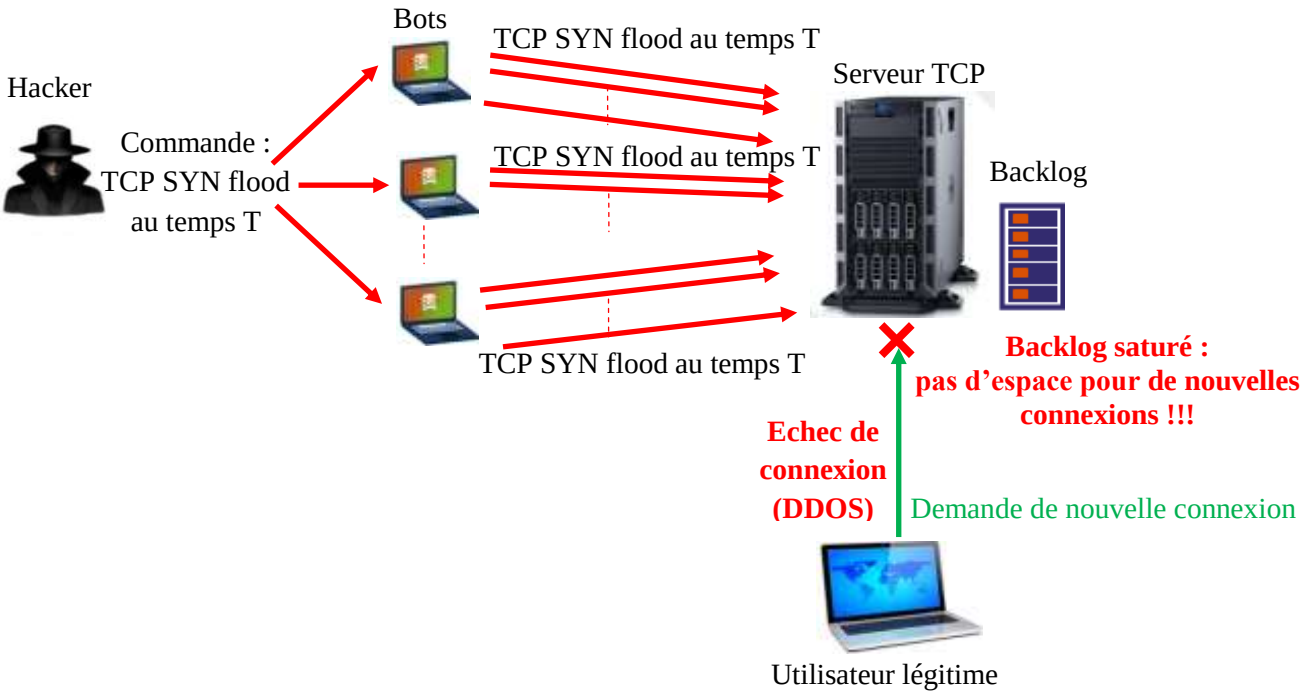

Figure 1.8: Attaque DDOS TCP SYN flood

1.4.2. Attaque UDP flood

Une attaque DOS UDP flood est crée lorsque l'attaquant envoie un datagramme UDP à un port aléatoire sur la victime ciblée. Si le port de destination est fermé, elle répondra par un message ICMP à l'adresse IP source usurpée, signalant l'erreur de port de destination inaccessible. Si suffisamment de datagrammes UDP sont envoyés vers des ports fermés (figure 1.9), la victime et même les hôtes sur le même segment seront également hors service en raison de la grande quantité de trafic engendré [12].

Un autre exemple d'une inondation UDP couramment utilisée est l'attaque par déni de service Chargen (figure 1.10). Dans ce scénario, l'attaquant connecte le service chargen de la première machine victime au service d'écho d'une deuxième victime. Le premier génère des caractères, tandis que le second retransmet simplement les données qu'il reçoit. Ainsi, l'attaquant usurpe l'adresse IP de la première victime et utilise son port 7 (Echo) pour envoyer le datagramme UDP au port 19 (chargen) sur la deuxième victime. Par conséquence, le déluge de datagrammes UDP provoquera une saturation des deux machines [16].

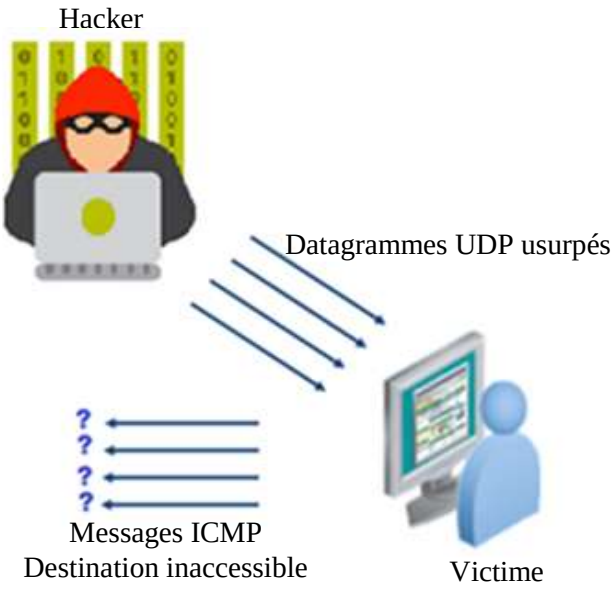

Figure 1.9 : Attaque DOS UDP flood

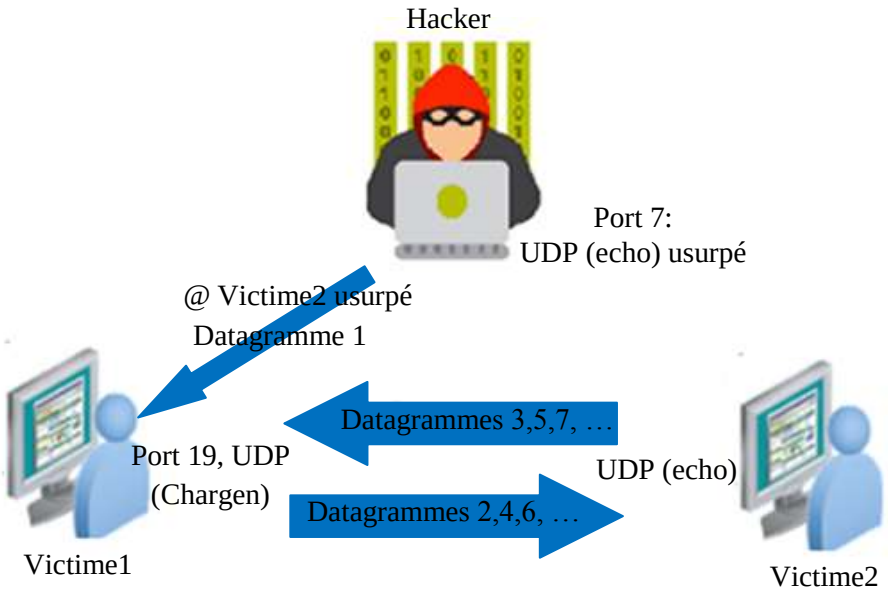

Figure 1.10 : Attaque DOS UDP Chargen

1.4.3. Attaque SMURF

L'architecture TCP/IP implémente le protocole ICMP pour transmettre les messages d'administration (par exemple, Traceroute et Ping) et les messages d'erreurs, tels que réseau /hôte/protocole/destination inaccessibles, durée de vie expirée, fragmentation requise…etc. Cependant, les messages ICMP sont fréquemment exploités par les hackers pour mener à terme plusieurs formes d'attaques DOS et DDOS (par exemple, ping of death, ping flood et SMURF) [17].

Les attaques par amplification ICMP, également baptisées attaques SMURF représentent un type pertinent d'attaques DDOS, qui peuvent être basées sur l'ICMPv4 tout comme sur la version ICMPv6, menaçant tant les réseaux IPv4 que l'environnement IPv6 [18]. En fait, les attaques SMURF utilisent les serveurs de diffusion pour submerger leurs victimes avec un flux élevé de messages ICMP ECHO-REPLY. Pour accomplir une telle attaque, l'attaquant commence par usurper l'adresse IP de la victime ciblée, puis envoie des messages ICMP ECHO-REQUEST au serveur de diffusion qui les achemine vers tous les hôtes de son domaine de diffusion (figure 1.11). Ces derniers répondent avec le message ICMP ECHO-REPLY à l'adresse IP de la victime. Etant donné qu'un domaine de diffusion peut inclure éventuellement des centaines d'hôtes, la victime sera inondée par un énorme flux de messages ECHO-REPLY qui épuisent ses ressources et la mettra hors service [1] [17].

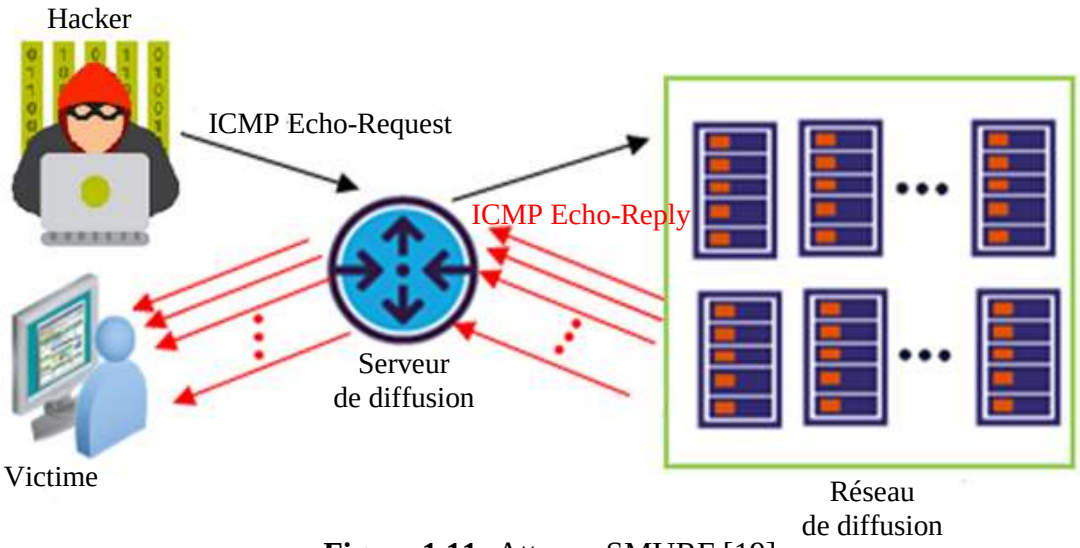

Figure 1.11: Attaque SMURF [19]

1.4.4. Attaques DOS et DDOS à base du protocole ICMPv6

Pour satisfaire les nouvelles applications de l'Internet, la version 6 du protocole IP (IPv6) a été conçue comme une solution adéquate pour dépasser la pénurie d'adresses IPv4. Elle offre de nombreuses améliorations par rapport au protocole IPv4, telles qu'un routage efficace, le traitement des paquets, la configuration automatique des adresses, la prise en charge du multicast et la connectivité peer-to-peer [20]. Les fonctionnalités de base de ces améliorations, à savoir la découverte, la sollicitation et l'annonce de voisin, la sollicitation et l'annonce de routeur, et celles d'autres protocoles IPv4 (par exemple, le protocole de résolution d'adresse ARP (Address Resolution Protocol) et le protocole de gestion de groupe Internet IGMP (Internet Group Management Protocol) ont été toutes associées à la version six du protocole ICMP (ICMPv6) [18]. En effet, le protocole ICMPv6 est considéré comme un

élément clé des réseaux IPv6. Pour assurer toutes ces tâches, le protocole ICMPv6 fournit de nombreux types de messages, mais sans aucun mécanisme de protection concernant leur génération, leur transmission et leur réception. Cette lacune de sécurité le rend vulnérable aux différents types de cyber-attaques, y compris plusieurs attaques DOS et DDOS, telles que les ping flood ICMPv6, les SMURF ICMPv6 et les RSMURF, la découverte de routeurs et la découverte de voisins [18] [21] [22].

- **Attaque par NS flood**

  Les messages de sollicitation de voisins (NS : Neighbor Solicitation) sont généralement utilisés par les hôtes clients pour effectuer la fonction de résolution d'adresse IPv6. Du fait qu'il n'y a pas de restriction, tous les hôtes, y compris les anormaux et les attaquants, peuvent exécuter cette procédure, et les destinataires de ces messages doivent répondre avec les messages d'annonce de voisin (NA : Neighbor Advertisement) annonçant ainsi leurs propres adresses IPv6 (figure 1.12). Or, lors d'une attaque par inondation de message NS (NS flood), l'hôte cible est surchargé et se met hors service lorsqu'il essaie de répondre à toutes ces demandes.

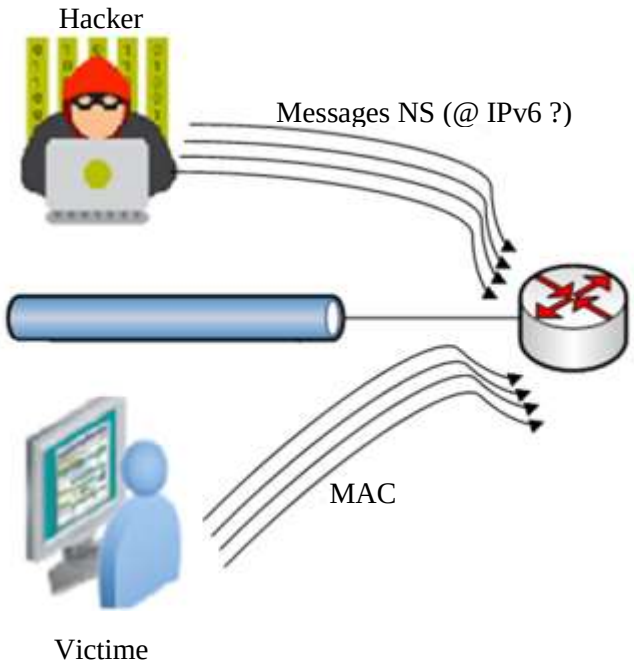

Figure 1.12: Attaques DOS à base de messages ICMPv6 NS (NS flood)

- **Attaque par NA flood**

Les messages NS sont utilisés par le protocole DAD (Duplication Address Detection) qui permet à chaque hôte d'effectuer le mécanisme d'auto-configuration et de vérifier si l'adresse IPv6 qu'il souhaite allouer n'est pas déjà utilisée par un autre hôte du réseau. Pour créer une attaque par NA (Neighbor Advertisement) flood, l'attaquant répond à tous les messages NS annonçant, régulièrement, qu'il a déjà l'adresse demandée (figure 1.13). Par conséquence, la machine victime n'arrivera jamais à avoir une adresse IPv6 et ne pourra donc pas effectuer de communications internes ou externes [18].

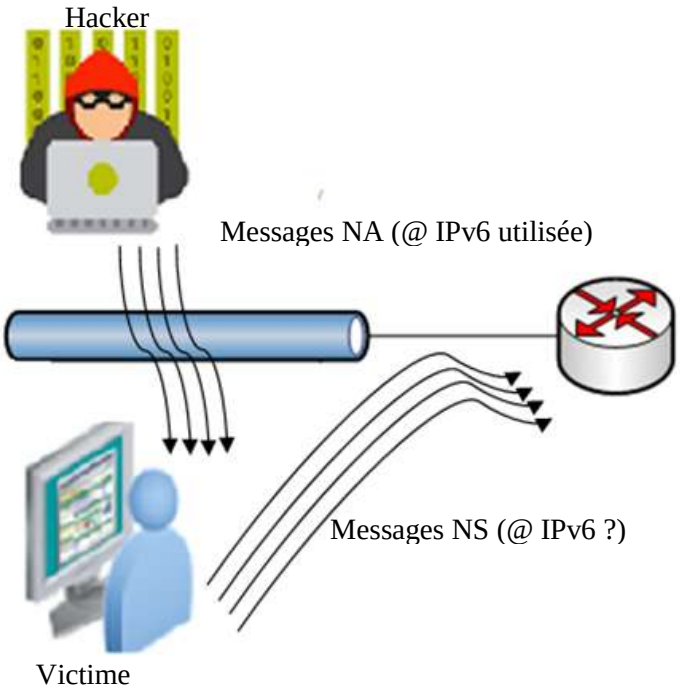

Figure 1.13: Attaques DOS à base de messages ICMPv6 NA (NA flood)

- **Attaque par RA flood**

L'attaquant usurpe l'adresse IP du routeur légitime et envoie des messages Router Advertisement (RA) avec le champ Router Lifetime mis sur 0. A la réception des messages usurpés, les hôtes victimes supprimeront le routeur par défaut de leur table de routage et ne seront plus en mesure d'établir des communications en dehors de leur réseau local (figure 1.14) [18].

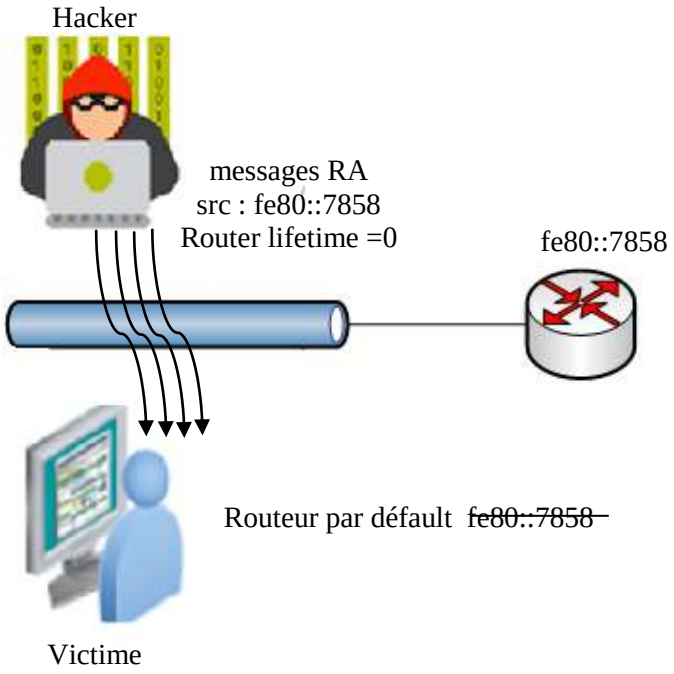

Figure 1.14: Attaques DOS à base de messages ICMPv6 RA (RA flood)

1.5. Détection des attaques DOS/DDOS

Les attaques DOS et DDOS prennent de diverses formes et constituent une menace extrêmement pernicieuse au sein du domaine de l'Internet. Malgré le nombre important de recherches effectuées au cours de la dernière décennie, la détection précoce et précise des attaques DOS et DDOS démure un vrai défi. Il n'est pas toujours facile de différencier les flux malveillants des flux légitimes [1]. En effet, le développement de techniques et de mécanismes pour une détection efficace et rapide de telles attaques continue d'être un domaine de recherche très actif. Deux grandes familles de techniques de détection se trouvent dans la littérature : la détection à base de signature et la détection à base d'anomalie.

- **Détection à base de signature (Signature-based detection)**

Les techniques de détection à base de signature analysent les activités des systèmes, recherchant des événements (ou des ensembles d'événements) qui correspondent à un modèle d'événements prédéfini. Cela implique l'analyse de signatures qui représentent un schéma d'attaque connu. Une signature peut être l'interprétation de séries de paquets ou d'une donnée contenue dans ces paquets. Elle peut également se manifester dans les enregistrements d'audit, les journaux ou les modifications des fichiers ou de la mémoire du système compromis.

Ainsi, Ces techniques parviennent à détecter les attaques DOS/DDOS connues. Cependant, tous les changements dans les attaques demeurent inaperçus par ces mécanismes de détection. De plus, comme ces techniques ne possèdent pas les nouvelles signatures, elles ne peuvent pas détecter les nouveaux types d'attaques qui n'ont pas été vus auparavant [10].

Par conséquence, elles doivent être régulièrement mises à jour avec des signatures de nouvelles attaques. La définition de la signature est une tâche critique. Si les signatures sont définies de manière lâche, plus d'attaques seront détectées mais avec plus de fausses alarmes. En revanche, si les signatures sont étroitement définies, cela réduira le nombre de fausses alarmes mais risque de rater plus d'attaques [21].

- **Détection à base d'anomalie (Anomaly-based detection)**

La détection des anomalies identifie toute déviation inacceptable par rapport à un comportement prévu. Elle suppose que les attaques sont différentes de l'activité normale et peuvent donc être détectées en identifiant ces différences. Les comportements attendus des utilisateurs, des hôtes ou des connexions réseau sont construites à l'avance. Les profils peuvent être créés manuellement ou automatiquement sur la base de données préalablement collectés sur une période de fonctionnement normal (supposée exempt d'attaques) [23].

Malheureusement, ces techniques nécessitent souvent une phase d'apprentissage et leurs performances sont très sensibles aux données d'apprentissage. Par conséquence, ils produisent souvent un grand nombre de fausses alarmes, de fait que les comportements normaux des utilisateurs et des systèmes peuvent varier considérablement. Malgré cette limitation, les chercheurs affirment que ces techniques sont capables de révéler de nouvelles formes d'attaques, contrairement aux techniques à base de signatures [24].

1.6. Conclusion

Dans ce chapitre nous avons fait le tour sur les attaques de dénie de service DOS et DDOS dans les réseaux IP. Nous avons vu que par définition, ces attaques ciblent principalement la disponibilité des actifs que se soient matériels ou logiciels. Elles sont en croissance permanente, et de nouvelles attaques apparaissent continuellement. Selon leur impact, souvent dévastateur, on peut les grouper dans deux grandes classes : d'épuisement des ressources et d'épuisement de la bande passante. Dans chacune des classes, on distingue différents types d'attaques qui exploitent différemment les failles et les limitations dans les protocoles des différentes couches. En particulier, nous avons présenté brièvement le principe

des attaques TCP SYN flood, UDP flood, SMURF et les attaques basées sur le protocole ICMPv6 (inondations par sollicitation et annoncent de voisin et par annonce de routeur). Pour les neutraliser, deux familles d'approches de détection sont principalement déployées : détection à base de signature et à base d'anomalie.

Vue leur performances prometteuses, les techniques de détection d'anomalies ont été largement étudiées par les chercheurs qui désirent mettre en place des solutions plus efficaces et plus appropriées. Ces techniques font l'objet d'une étude pertinente dans le chapitre suivant.

Chapitre2

Mitigation des cyber-attaques par les techniques de détection d'anomalies

Dans ce chapitre

- La détection d'anomalies
- Types d'anomalies dans les réseaux IP
- Modes de détection d'anomalies
- Mesures d'évaluation
- Techniques de détection d'anomalies et les cyber-attaques

Chapitre 2

Mitigation des cyber-attaques par les techniques de détection d'anomalies

Dans le processus de surveillance des systèmes, la tâche de détection des anomalies constitue une opération critique. Elle s'adressait au problème d'identification de tous types de défaillances qui peuvent changer, significativement, le comportement du système (souvent indésirable). Le défi consiste à pouvoir détecter rapidement les anomalies, même en temps réel, pour garder le système fiable et à l'abri de leurs conséquences potentiellement catastrophiques [25].

La détection d'anomalies a été largement appliquée dans plusieurs domaines, notamment des réseaux informatiques (systèmes de détection d'intrusion, sécurité, surveillance du trafic...), détection de fraude (banques, carte de crédits...), les systèmes de santé, traitement d'image, la reconnaissance vocale et les systèmes CPS (Cyber Physical Systems : réseaux de distribution d'électricité, d'eau et de gaz) [26] [27] [28].

Compte tenu de leurs capacités prometteuses, les systèmes de détection basés sur la détection d'anomalies constituent actuellement l'un des principaux axes de recherche et de développement dans le domaine de la détection des intrusions et des cyber-attaques. En fait, divers systèmes A-NIDS (Anomaly-based Network Intrusion Detection System) sont de plus en plus disponibles, et de nombreux nouveaux systèmes sont à l'étude. Toutefois, le sujet est loin d'être mature et des questions clés restent à résoudre avant que le déploiement à grande échelle des plates-formes A-NIDS puisse être réalisable.

Dans ce chapitre nous présentons un état de l'art sur la détection d'anomalies et son application dans la détection des attaques et d'intrusions dans les réseaux IP. On commencera par la définition du concept de détection d'anomalie et son principe de fonctionnement. On citera ensuite les différents types d'anomalies dans un réseau, suivi par les modes principaux de détection et les métriques qui sont souvent utilisées pour évaluer leurs performances. On terminera par une présentation générale des travaux menés pour le développement des techniques et des systèmes de détection.

2.1. Définition de la détection d'anomalies

La détection d'anomalies se réfère au problème de recherche de modèles dans des données qui ne sont pas conformes au comportement normal attendu. Ces schémas non conformes sont souvent appelés anomalies.

La figure 2.1 illustre les anomalies d'un simple ensemble de données en deux dimensions. Les données ont deux régions normales, N1 et N2, puisque la plupart des observations se trouvent dans ces deux régions. Les points qui sont suffisamment éloignés des régions, par exemple les points o1 et o2, et les points dans la région O3, sont des anomalies [26] [29].

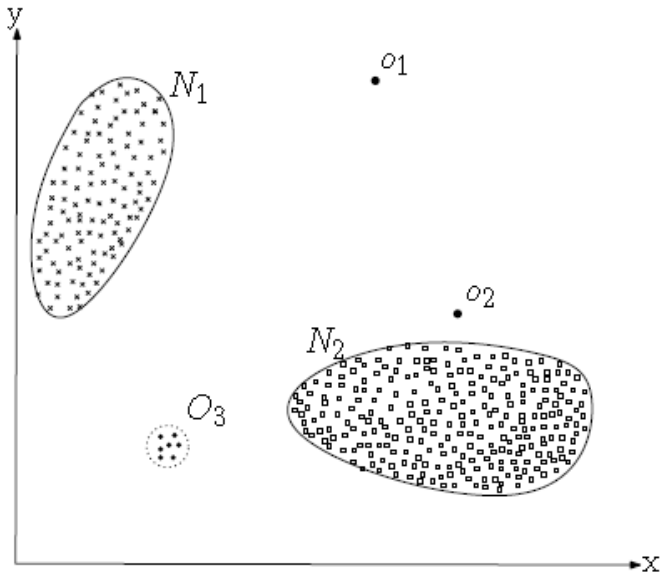

Figure 2.1: Exemples d'anomalies dans des données à deux dimensions [26]

L'importance de la détection des anomalies est due au fait que les anomalies dans les données se traduisent par des informations exploitables significatives (et souvent critiques) dans une grande variété de domaines d'application.

2.2. Types d'anomalies dans les réseaux IP

Les anomalies réseau se réfèrent généralement à des circonstances où les opérations réseau divergent de leur comportement normal. Les anomalies réseau peuvent survenir en raison de diverses causes telles que les équipements réseau défectueux, l'état de congestion, les attaques de déni de service et les différentes formes d'intrusions. Ces événements

anormaux perturbent la prestation normale des services réseau ainsi que le comportement normal de certaines données réseau mesurables [2].

La définition du comportement normal d'un réseau pour des données (ou paramètres) mesurées dépend de plusieurs facteurs spécifiques au réseau tels que la dynamique du réseau en terme de volume de trafic, le type de données réseau disponibles et les types d'applications en cours d'exécution sur le réseau. Par conséquence, un événement ou un objet est qualifié comme anormal si son degré de déviation par rapport au profil ou au comportement du système, spécifié par le modèle de normalité, est assez élevé.

Les anomalies réseau peuvent être classées en deux catégories [2] [27] :

(i) La première catégorie est liée aux défaillances réseau et aux problèmes de performances. Les exemples typiques d'anomalies des performances réseau sont les défaillances des serveurs de fichiers, la pagination sur le réseau, les rafales de diffusion et l'état de congestion;

(ii) La deuxième catégorie majeure d'anomalies réseau est liée aux problèmes de sécurité. Les attaques par déni de service et les intrusions de réseau sont des exemples de telles anomalies.

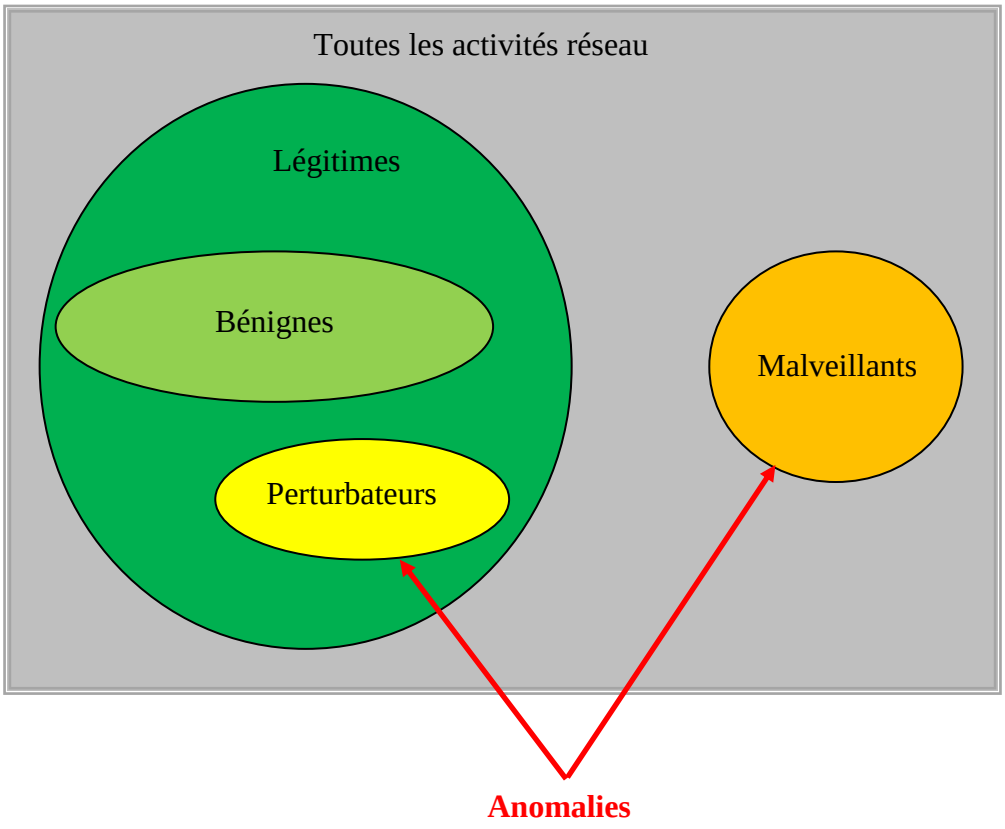

Figure 2.2: Types des activités réseau [30]

2.3. Modes de détection d'anomalies

Parmi des techniques de détection des anomalies, il existe différentes techniques ont été largement exploitées pour établir la distribution normale du trafic, c'est principalement l'apprentissage supervisé et l'apprentissage non-supervisé. Nous présentons ici les différentes techniques d'apprentissage qui sont couramment utilisées et nous énumérons leurs points forts ainsi que leurs principales limitations [31].

2.3.1. Apprentissage supervisé

L'apprentissage supervisé est la branche la plus courante des techniques de détection des anomalies. Lorsque le processus est basé sur la disponibilité de données d'apprentissage qui ont été classées et étiquetés en groupes distincts, et il est incontestable de faire la différence entre les comportements normaux et ceux anormaux. Le processus de détection d'anomalies implique une simple comparaison des enregistrements de test avec le modèle d'apprentissage qui représente l'état normal des données [26].

Deux problèmes majeurs se posent dans la détection supervisée des anomalies. Premièrement, les cas anormaux sont beaucoup moins nombreux que les cas normaux dans les données d'apprentissage. Les enjeux liés à la répartition déséquilibrée des classes ont été abordés dans la littérature de datamining et machine learning [32]. Deuxièmement, il est généralement difficile d'obtenir des étiquettes précises et représentatives, en particulier pour la classe d'anomalie. Un certain nombre de techniques injectent des anomalies artificielles dans un ensemble de données normal pour obtenir de bases de données d'apprentissage étiquetées [33].

2.3.2. Apprentissage semi-supervisé

Les techniques semi-supervisées supposent que les données d'apprentissage ont étiqueté des instances uniquement pour la classe normale. Comme ils n'ont pas besoin d'étiquettes pour la classe d'anomalies, ils peuvent être plus facilement utilisés que les techniques supervisées. L'approche typique utilisée dans de telles techniques est de construire un modèle pour la classe qui correspond au comportement normal, et de l'utiliser pour identifier les anomalies dans les données de test [26].

L'apprentissage semi-supervisé fonctionne comme le supervisé, mais dans ce cas, le référence pour un comportement attendu ou un enregistrement standard est déjà défini, et les

anomalies sont tout le reste qui se trouve en dehors de la classe. En raison de cet attribut, ce type d'apprentissage est utilisé dans de nombreux domaines pour leur convenance dans les situations où la classe d'anomalies n'est pas adéquatement connue [26].

2.3.3. Apprentissage non-supervisé

Les techniques non-supervisées ne nécessitent pas de données d'apprentissage et sont donc potentiellement les plus largement applicables. Les techniques de cette catégorie sont basées sur l'hypothèse que les cas normaux sont beaucoup plus fréquents que les anomalies dans les données de test [34] [35]. Lorsque cette hypothèse n'est pas satisfaite, telles techniques souffrent de taux élevés de fausses alarmes. De nombreuses techniques semi-supervisées peuvent être adaptées pour fonctionner en mode non-supervisé en utilisant un échantillon de l'ensemble de données non étiquetés comme données d'apprentissage [36]. Cette adaptation suppose que les données de test contiennent très peu d'anomalies et que le modèle construit au cours de l'apprentissage est robuste à ces quelques anomalies.

2.4. Mesures d'évaluation

Une évaluation d'une technique ou d'un système de détection d'intrusion réseau est instantanée dans le temps. Au fil du temps, de nouvelles vulnérabilités peuvent être découvertes et les évaluations actuelles peuvent devenir non pertinentes [31]. Par ailleurs, l'objectif d'un IDS est de produire autant de TP (True Positive) et de TN (True Negative) que possible tout en essayant de réduire le nombre de FP (False Positive) et de FN (False Negative). Nous discutons, ci-dessous, diverses mesures utilisées pour évaluer les méthodes et les systèmes de détection des intrusions et des cyber-attaques réseau.

- **Accuracy:**

Accuracy est une métrique qui reflète le fonctionnement correct d'un IDS, mesure le pourcentage de détection et de défaillance ainsi que le nombre de fausses alarmes que le système produit [37]. Ainsi, si un système a une précision de 80%, cela signifie qu'il classe correctement 80 instances sur 100 à leurs classes réelles [31]. Bien qu'il existe une grande diversité d'attaques dans la détection des intrusions, l'objectif principal est que le système soit en mesure de détecter une attaque correctement. De la pratique, on peut facilement conclure que le pourcentage de données anormaux est beaucoup plus faible que celui des normaux

[34]. Par conséquence, les intrusions sont plus difficiles à détecter que le trafic normal, ce qui entraîne des fausses alarmes excessives comme le plus grand problème auquel sont confrontés les IDS.

- **TP, FN, TN, FP**

Lorsqu'une instance anormale de test (P) est détectée comme anormale (P) par le détecteur, elle est comptée comme True Positive (TP); si elle est détectée comme normal (N), elle est comptée comme False Negative (FN). D'autre part, si une instance normale (N) est détectée comme normale (N), elle est connue sous le nom de True Negative (TN), alors qu'il s'agit d'un False Positive (FP) si elle est détectée comme anormale (P) [38].

- **True Positive Rate (TPR), False Positive Rate (FPR), True Negative Rate (TNR) et FNR (False Negative Rate) [31]**

Le TPR constitue la proportion de cas anormaux classés correctement par rapport au nombre total de cas anormaux présents dans les données de test. Il est également connu sous le nom de sensibilité.

Le FPR est la proportion de cas normaux mal classés comme anormaux par rapport au nombre total de cas normaux contenus dans les données de test.

Le taux négatif réel (TNR) est également appelé spécificité.

TPR, FPR, TNR et FNR peuvent être définis pour la classe normale.

Detected Class	True Class: P	True Class: N
P	True Positive TP	False Positive FP
N	False negative FN	True Negative TN

Figure 2.3: Matrice de confusion [31]

Soient :

$Pos = TP + FN$ le nombre total de positives

$Neg = FP + TN$ le nombre total de négatives

Les métriques ci-dessus peuvent être calculées comme :

$$Accuracy = \frac{TP + TN}{Pos + Neg} \qquad (2.1)$$

$$FPR = \frac{FP}{Neg} = \frac{FP}{FP + TN} \quad (2.2)$$

$$TPR = \frac{TP}{Pos} = \frac{TP}{TP + FN} \quad (2.3)$$

$$TNR = \frac{TN}{Neg} = \frac{TN}{FP + TN} = 1 - FPR \quad (2.4)$$

$$FNR = \frac{FN}{Pos} = \frac{FN}{TP + FN} = 1 - TPR \quad (2.5)$$

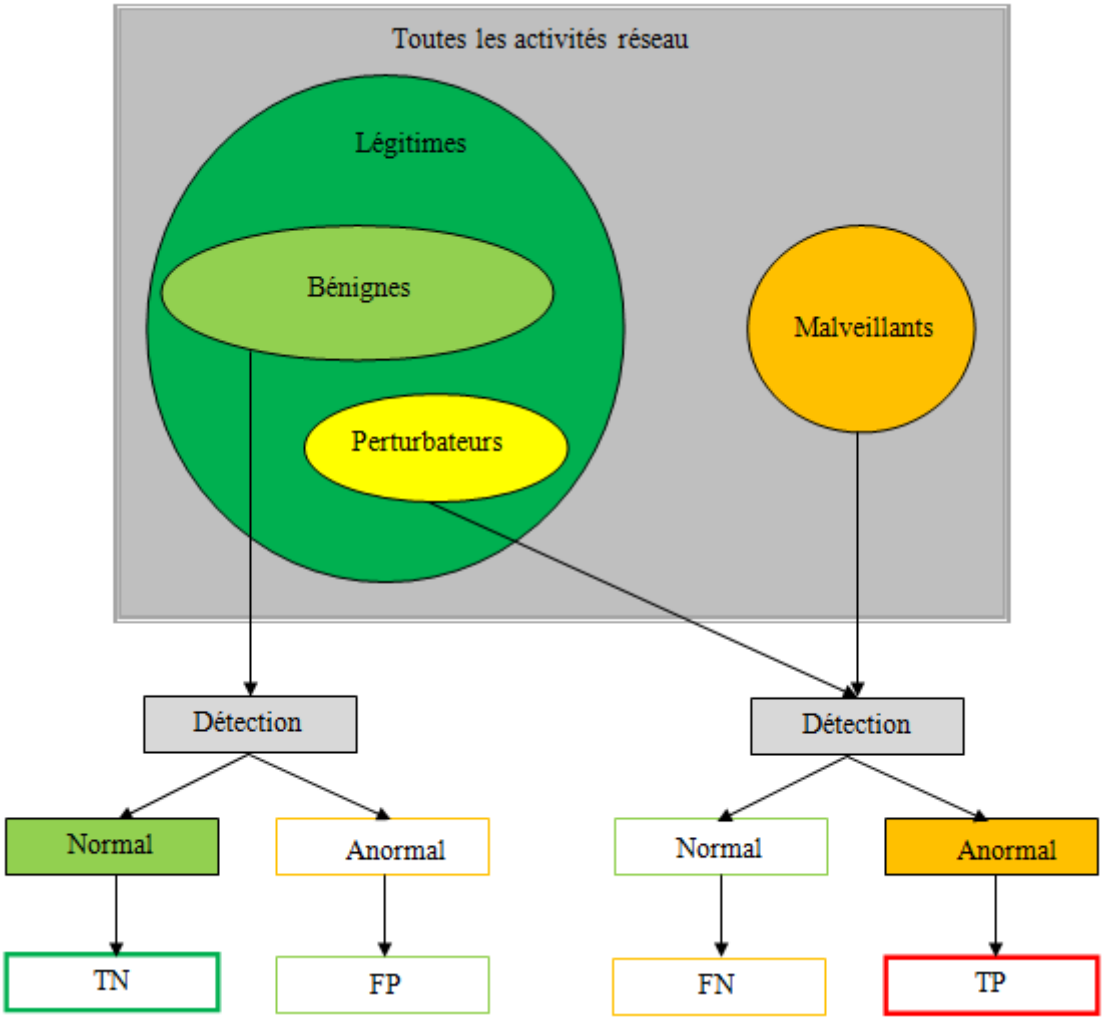

Figure 2.4: Types des activités réseau et mesures d'évaluation [30]

2.5. Architecture générale d'un A-NIDS

Bien qu'ils existent différentes approches A-NIDS, le développement d'une architecture A-NIDS efficace est toujours un véritable challenge pour les acteurs du domaine. Comme illustré dans la figure 2.5, un A-NIDS typique se compose de quatre éléments

principaux: source de données, module de prétraitement des données, engin de décision DE (Decision Engine) et réponses de sécurité [31] [23] [39].

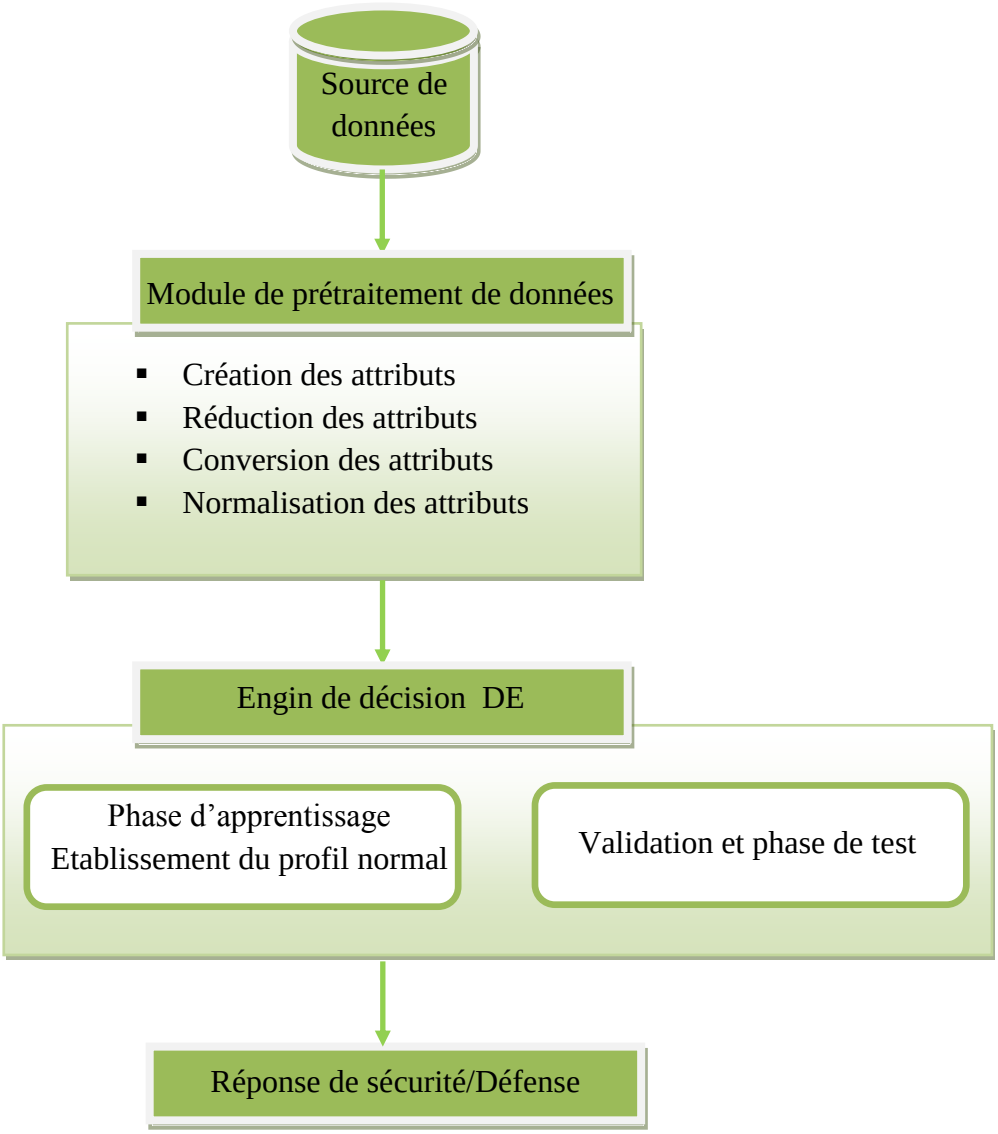

Figure 2.5: Architecture générique d'un A-NIDS

2.5.1. Source de données

La source de données est une composante majeure de tout A-NIDS pour l'évaluation des performances des méthodes DE, en raison de la difficulté d'étiqueter les activités légitimes et d'attaques dans le trafic réseau en ligne [40]. Les sources de données réseau peuvent être collectées à partir d'un trafic temps réel ou d'une base de données hors ligne qui comprend une grande variété d'événements normaux et malveillants.

2.5.2. Module de prétraitement des données

Le prétraitement des données est une étape importante dans la théorie de l'apprentissage. En fait, les outils de collecte de données sont souvent peu contrôlés et qui entraînent des données brutes et de caractéristiques non pertinents ou dupliquées; les données extraites du trafic réseau n'échappent pas de la règle et incluent également différent types de paramètres. Le module de prétraitement filtre ces données en supprimant les informations redondantes, bruyantes ou non pertinentes, ce qui permet d'améliorer les performances des approches DE pour détecter les comportements d'attaques. Le prétraitement des données réseau implique la création, la réduction, la conversion et la normalisation des attributs.

2.5.3. Engin de décision DE

Le module DE d'un A-NIDS est clairement un aspect essentiel dans la conception d'un système efficace pour découvrir les activités intrusives en temps réel. Les approches DE peuvent être classées en plusieurs catégories [27] [41] [42], telles qu'elles sont représentées dans la figure 2.6

2.5.4. Réponses de sécurité

Les réponses de défense ou de sécurité sont des mesures prises par le système contre les événements malveillants détectés. Plus précisément, Elles ont la capacité d'identifier une activité donnée comme une attaque, puis l'administrateur du système devrait prendre une mesure pour arrêter l'activité malveillante [42]. Il existe deux types de réponses : passive et active.

- **Réponse passive :** est prise par un administrateur humain lorsqu'un IDS identifie un événement malveillant. Ce processus se produit normalement après la collecte et la corrélation des traces par l'administrateur, lorsqu'un comportement anormal est détecté et qu'une alerte est déclenchée. La forme populaire d'une alarme est une fenêtre contextuelle ou une alerte à l'écran. Il peut être affiché sur la console IDS telle que la console d'alerte SNORT. Il existe des traps et des messages SNMP (Simple Network Management Protocol) qui créent des alertes et des rapports à l'administrateur du réseau pour prendre les mesures adéquates.
- **Réponse active :** est également appelée système de prévention des intrusions (IPS : Intrusion Prevention System), qui est une action immédiate et automatique prise

lorsque des événements malveillants sont détectés par l'exécution d'une action de script prédéfinie. Elle permet d'arrêter la progression des attaques en bloquant leurs adresses IP et leurs ports, en modifiant l'ACL (Access Control List), en réinitialisant le protocole TCP pour la fin des connexions et/ou en reconfigurant les pare-feu et les routeurs.

2.6. Techniques de détection d'anomalies utilisées dans les IDS

La classification des techniques et des systèmes de détection des anomalies réseau que nous adoptons est indiquée à la figure 2.6. Ce schéma est basé sur la nature des algorithmes utilisés [31] [42].

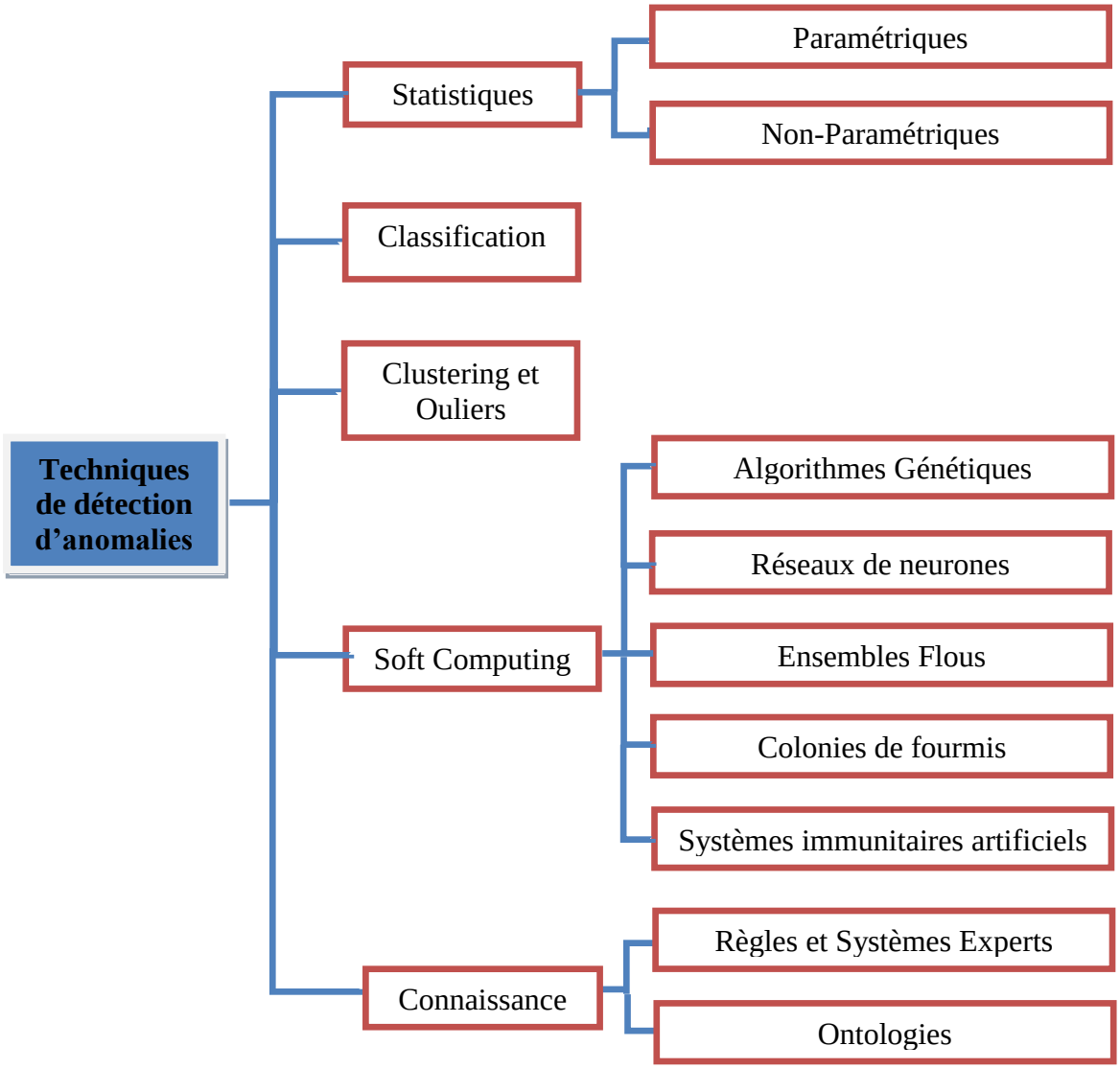

Figure 2.6: Classification des techniques de détection d'anomalies [31]

2.6.1. Techniques et systèmes statistiques

Dans les techniques statistiques, l'activité du trafic réseau est capturée et un profil représentant son comportement stochastique est crée. Ce profil est basé sur des mesures telles que le débit de trafic, le nombre de paquets pour chaque protocole, le taux de connexions, le nombre d'adresses IP différentes, etc. Deux ensembles de données du trafic réseau sont pris en compte au cours du processus de détection des anomalies : l'un correspond au profil actuellement observé, et l'autre est pour le profil statistique précédemment construit. Au fur et à mesure que les événements réseau se produisent, le profil actuel est déterminé et un score d'anomalies est estimé par comparaison des deux comportements. Le score indique normalement le degré d'irrégularité pour un événement spécifique, de sorte que le système de détection d'intrusion signalera l'apparition d'une anomalie lorsque le score dépasse un certain seuil.

Des techniques paramétriques et non-paramétriques ont été appliquées pour concevoir des modèles statistiques de détection d'anomalies. Alors que les techniques paramétriques supposent la connaissance de la distribution sous-jacente et estiment les paramètres à partir des données [43], les techniques non-paramétriques n'assument généralement pas la connaissance de la distribution sous-jacente [44].

Zhang et al. [45] ont proposé le HIDE (Hierarchical Intrusion DEtection), un système de détection d'intrusion réseau basé sur des anomalies, qui utilise des modèles statistiques et des classificateurs de réseaux neuronaux pour détecter les intrusions. HIDE est un système distribué, qui se compose de plusieurs niveaux avec chaque niveau contenant plusieurs agents de détection d'intrusion (IDA : Intrusion detection Agent). Les IDA sont des composants IDS qui surveillent les activités d'un hôte ou d'un réseau.

Chen et al. [46] combinaient le coefficient maximaux d'information (MIC : Maximum Information coefficient) et l'analyse par composante principal multiéchelle (MSPCA : Multi-Scale Principal Component Analysis) pour détecter les anomalies dans le trafic réseau. MSPCA est appliquée pour détecter les anomalies en fonction des caractéristiques sélectionnées à l'aide de MIC, qui comprend, mais ne se limite pas au volume de trafic. La proposition des auteurs est d'agréger les paquets réseau acquis à l'aide de TCPDUMP dans les flux réseau, et d'appliquer la méthode présentée aux données colléctés.

Altwaijry [47] a développé un A-NIDS BN (Bayesian Network) naïf utilisant le PCA (Principal Composant Analysis). Un BN est une distribution de probabilité graphique pour prendre des décisions concernant des données incertaines [48]. Dans cette approche, le PCA

calcule les caractéristiques les mieux classées et utilise les caractéristiques sélectionnées et leurs composants comme poids pour améliorer la technique traditionnelle BN naïve.

Gruhl et al. [49] ont mis en place un A-NIDS basé sur la combinaison de mécanismes de modélisation de la densité paramétrique et non-paramétrique en deux étapes. Premièrement, les échantillons malveillants ont été reconnus à l'aide du GMM (Gaussian Mixture Model) puis regroupés dans une mesure non-paramétrique dans la deuxième étape. Alors qu'un groupe s'étendait jusqu'à une taille adéquate, une procédure a été identifiée, transformée en mesure paramétrique et ajoutée au GMM établi. Cette technique a été évaluée à l'aide de la base de données KDD99 (Knowledge Discovery and Data Mining 99) et leurs résultats reflétaient une accuracy de détection élevée et un faible FPR. Toutefois, ils exigeraient l'utilisation de l'inférence Bayésienne soit ajustée en fonction de leur application pour être efficace dans un réseau réel.

Shahriar et al. [50] ont appliqué l'entropie pour détecter les requêtes SQL (Structured Query Language) vulnérables dans les applications web PHP (Hypertext Preprocessor). Plus tard, ils ont exploré une métrique de théorie de l'information basée sur la mesure de la dissimilarité KLD pour détecter les attaques XSS (Cross-site scripting) dans les applications web [51]. Les mesures KLD ont été proposées, aussi pour détecter les logiciels malwares Android reconditionnés [52]. Ozonat et al. [53] détectaient des anomalies dans le comportement de la métrique de performance dans les services web distribués à grande échelle en appliquant des mesures théoriques de l'information.

Les auteurs dans [54] ont appliqué la distance de Hellinger (HD : Hellinger Distance) sur la structure de données de Sketch, afin de détecter la divergence entre les distributions actuelles et précédentes du nombre de requêtes SIP INVIT (Session Initiation Protocol Invitation). En fait, la HD doit être proche de zéro lorsque les distributions de probabilité sont similaires, et elle augmente si les distributions divergent (par exemple en présence des attaques Invitation flood). En outre, ils ont utilisé un seuil dynamique dans leur analyse expérimentale.

2.6.2. Techniques et systèmes de classification

La classification consiste à identifier à quelle catégorie appartient une nouvelle observation, en se basant sur un ensemble de données d'apprentissage contenant des observations dont leur appartenance à une catégorie est bien connue.

Ainsi, les techniques de classification sont basées sur l'établissement d'un modèle explicite ou implicite qui permet de catégoriser les modèles de trafic réseau en plusieurs classes [55] [56] [57]. Une caractéristique singulière de ces techniques est qu'elles requirent de données étiquetées pour établir le modèle comportemental. Telle procédure impose des exigences élevées en termes de ressources [58]. Dans de nombreux cas, l'applicabilité des principes d'apprentissage automatique (machine learning) tels que la classification coïncide avec celle des techniques statistiques, bien que l'ancienne technique soit axée sur la construction d'un modèle qui améliore ses performances sur la base des résultats précédents [23].

Les techniques de classification les plus populaires appliquées pour les A-NIDS sont la SVM (Support Vector Machine), le KNN (K-Nearest Neighbour), ainsi que les NN (Neural Networks) [39].

Wagner et al [59] ont utilisé des classificateurs d'une classe qui peuvent détecter de nouvelles anomalies, c'est-à-dire des points de données qui n'appartiennent pas à la classe appris. En particulier, ils ont exploité un classificateur SVM d'une classe proposé par Schölkopf et al. [60]. Dans tel classificateur, les données d'apprentissage sont présumées appartenir à une seule classe, et l'objectif d'apprentissage est de déterminer une fonction positive lorsqu'elle est appliquée aux points de la limite circonscrite autour des points d'apprentissage et négative à l'extérieur. C'est aussi ce qu'on appelle la classification semi-supervisée. Un tel classificateur SVM peut être utilisé pour identifier les valeurs aberrantes et les anomalies. Les auteurs ont développé une fonction de noyau spéciale qui projette les points de données vers une dimension supérieure avant la classification. Leur fonction de noyau prend en considération les propriétés des données Netflow et permet de déterminer la similitude entre deux fenêtres des enregistrements de flux IP.

De même, Horng et al. [61] ont proposé un A-NIDS qui comprenait un clustering hiérarchique et un SVM afin de réduire le temps de traitement de la phase d'apprentissage et d'améliorer le taux de détection

Bhuyan et al. [62] ont mis en place un A-NIDS basé sur les KNN qui crée un profil réseau normal et traite toute déviation par rapport à celui-ci comme une attaque sans demandé aucune adaptation des paramètres à l'étape d'apprentissage. La technique KNN a été utilisée pour concevoir un A-NIDS fiable (DIDS : Dependable NIDS) basé sur l'étrangeté et les mesures d'isolement de ses fonctions potentielles qui pourraient identifier efficacement les attaques réseau.

2.6.3. Techniques et systèmes à base de Clustering et Outliers

Le clustering est la tâche d'attribuer un ensemble d'objets dans des groupes appelés clusters afin que les objets du même cluster soient plus semblables en quelque sorte les uns aux autres qu'à ceux d'autres clusters. Le clustering est utilisé dans l'exploration de données (datamining).

Bhuyan et al. [63] ont conçu un A-NIDS à base d'outlier pour les réseaux dont les données légitimes ont été regroupées à l'aide d'une technique k-means, puis un point de référence est calculé pour chaque cluster, ces points étant classés comme des attaques s'ils étaient inférieurs à une certaine valeur seuil. De plus, dans [64], Bhuyan et al. ont proposé un A-NIDS pour les grandes bases de données réseau utilisant des techniques de clustering par arbres et d'ensembles pour améliorer l'accuracy dans un environnement de réseau réel.

Les vers (Worms) sont souvent assez intelligents pour cacher leurs activités et échapper à la détection par les IDS. Zhuang et al. [65] proposent une méthode appelée PAIDS (Proximity-Assisted IDS) pour identifier les nouveaux vers lorsqu'ils commencent à se propager. PAIDS fonctionne différemment des autres IDS et a été conçu pour travailler en collaboration avec les IDS existants, tels que les IDS basés sur des anomalies, pour améliorer les performances. L'objectif des concepteurs de PAIDS est d'identifier les vers nouveaux et intelligents qui se propagent rapidement et de contrecarrer leur propagation, d'autant plus que le ver commence à peine à se propager. L'approche est basée principalement sur l'observation que pendant la phase de démarrage d'un nouveau ver, les hôtes infectés sont regroupés en termes de géographie, d'adresse IP et, peut-être, même de DNS utilisés.

Bhuyan et al. [64] présentent une méthode de détection d'anomalies réseau non-supervisée pour les grandes bases de données. Elle exploite le clustering à base d'arbres sous-espace et une technique d'étiquetage de cluster basée sur l'ensemble pour obtenir un meilleur taux de détection sur les données de trafic réseau réel pour la détection d'attaques connues ainsi que des attaques inconnues.

Otey et al. [66] élaborent une mesure de distance pour les données contenant un mélange d'attributs catégoriques et continus et les utilisent pour la détection d'anomalies à base d'outlier. Ils définissent un score d'anomalies qui peut être utilisé pour identifier les valeurs des outliers dans l'espace d'attributs mixtes en tenant compte des dépendances entre les attributs de différents types. Leur fonction de score d'anomalies est basée sur un modèle global des données qui peut être construit en combinant des modèles locaux construits indépendamment à chaque nœud. Ils ont développé un algorithme d'approximation à une

passe pour la détection des anomalies qui fonctionne efficacement dans les environnements de détection distribués avec très peu de perte d'accuracy de détection. Chaque nœud calcule ses propres valeurs d'outliers et la communication inter-nœuds nécessaire pour calculer les valeurs globales d'outliers n'est pas significative.

2.6.4. Techniques et systèmes du Soft Computing

Les techniques du Soft Computing conviennent à la détection d'anomalies réseau, car souvent on ne peut pas trouver de solutions exactes. Le Soft Computing englobe généralement des méthodes telles que les algorithmes génétiques GA (Genetic Algorithm), les réseaux de neurones NN, les ensembles flous, les algorithmes de colonies de fourmis (Ant Colony) et les systèmes immunitaires artificiels (AIS : Artificial immune system).

Khan [67] a utilisé des algorithmes génétiques pour élaborer des règles de détection des intrusions réseau. Un chromosome d'un individu contient des gènes correspondant à des attributs tels que le service, les indicateurs, connectés ou non, et les tentatives de super-utilisateur. Il conclut que les attaques qui sont fréquentes peuvent être détectées plus précisément par rapport à des attributs rares.

Tajbakhsh et al. [68] ont décrit une nouvelle méthode pour construire des classificateurs en utilisant des règles d'association floues et l'utiliser pour la détection des intrusions réseau. Les ensembles de règles d'association floues sont utilisés pour décrire différentes classes : normales et anormales. Il s'agit de règles d'association de classe où les conséquences sont des classes spécifiées. Elles sont induites à l'aide d'échantillons d'apprentissage normaux. Un échantillon de test est classé comme normal si la compatibilité de l'ensemble de règles générées est supérieure à un certain seuil; ceux qui ont une compatibilité plus faible sont considérés comme anormaux. Les auteurs proposent également une nouvelle méthode pour accélérer l'algorithme d'induction de règles en réduisant les éléments des règles extraites.

Les systèmes immunitaires artificiels AIS représentent une méthode de calcul inspirée des principes du système immunitaire humain. Le système immunitaire humain est habile à effectuer la détection des anomalies. Visconti et Tahayori [69] ont présenté un AIS basé sur la possibilité de détecter le comportement anormal individuel. Il surveille le système en analysant l'ensemble des paramètres pour fournir des informations générales sur son état. Le paradigme de jeu flou de type 2 d'intervalle est utilisé pour générer dynamiquement l'état du système.

2.6.5. Techniques et systèmes à base de connaissance (Knowledge)

Dans les méthodes basées sur la connaissance, les événements réseau ou hôte sont vérifiés en fonction des règles ou des schémas d'attaque prédéfinis. L'objectif est de représenter les attaques connues d'une manière généralisée afin que la manipulation des événements réels devienne plus facile.

Des exemples de techniques fondées sur le knowledge sont des systèmes experts, des règles, des ontologies, des analyses logiques et la transition par l'état [70] - [73]. Les approches basées sur des règles modélisent les connaissances recueillies sur les événements réseau suspects qui permettent la navigation des données de trafic réseau pour trouver des preuves de vulnérabilités existantes [74]. Un système d'expert comprend des règles qui définissent les événements d'attaques par lesquels les données de trafic réseau sont transformées en modèles, en fonction de leur poids relatif dans le système et un moteur d'inférence correspond aux règles prédéfinis avec l'état actuel du système pour détecter les activités d'attaques [39]. Ces approches ont été largement appliquées pour détecter les événements réseau suspects, tandis que les signatures d'intrusion basées sur l'ontologie et la logique sont basées sur une structure logique en intégrant les contraintes et les caractéristiques statistiques des données de trafic réseau [31].

Ces techniques recherchent des scénarios d'attaques connues, en essayant de faire correspondre avec des représentations d'attaques prédéterminées. La recherche commence comme d'autres techniques de détection d'intrusion, avec un manque total de connaissances. La correspondance ultérieure des activités contre une attaque connue permet d'acquérir des connaissances et d'entrer dans une région avec une plus grande confiance. Enfin, il peut être démontré qu'un événement ou une activité a atteint le score maximal d'anomalies [31].

L'outil SNORT est l'un des IDS populaires basé sur des règles et open source. Ses règles reconnaissent les paquets réseau malveillants en faisant correspondre le paquet actuel aux règles prédéfinies et ne peuvent pas détecter les attaques de jour zéro, mais produisent un FPR élevé en raison de sa méthodologie d'identification des signatures d'attaque [75]. A l'heure actuelle, SNORT comporte un nombre important de règles qui sont habituellement mises à jour par les utilisateurs [76]. L'outil des réseaux de Petri [77] a été conçu comme IDS à base de connaissance qui se compose de graphes bipartites dirigés et de réseaux de petri colorés (CPN : Colored Petri Nets) représentant les signatures d'intrusions. Cet outil a été utilisé pour développer l'outil Intrusion Detection In Our Time (IDIOT) pour détecter les événements indésirables [78]. Bien qu'il puisse facilement représenter de petites données

réseau et aide à discriminer les attaques connues, son processus d'appariement d'une signature d'attaque avec des règles prédéfinies est très difficile à exécuter dans des environnements réseaux réels et prend beaucoup de temps de traitement.

Duffield et al. [79] ont utilisé l'algorithme d'apprentissage automatique appelé Adaboost [80] pour traduire les signatures de niveau paquet et travailler avec des statistiques de niveau flux. L'algorithme est utilisé pour corréler le paquet et les informations de flux. En particulier, les auteurs associent les alarmes réseau au niveau des paquets à un vecteur de fonctionnalités qu'ils créent à partir d'enregistrements de flux sur le même trafic. Ils créent un ensemble de règles utilisant des informations de flux avec des fonctionnalités similaires à celles utilisées dans les règles SNORT. Ils ont ajouté également des fonctionnalités numériques telles que le nombre de paquets d'un type spécifique qui circulent dans une certaine période de temps. En fin, ils ont établit Adaboost sur des traces de flux et de paquets simultanés. Ils ont évalué le système en utilisant des données de trafic réseau en temps réel et montré que leurs performances sont comparables à celles de SNORT avec les données de flux.

Hung et al. [81] ont servi des ontologies comme un moyen de décrire la connaissance d'un domaine, exprimant le système de détection d'intrusion beaucoup plus en termes de domaine des utilisateurs finaux. Les ontologies sont utilisées comme outil de modélisation conceptuelle permettant à une personne non experte de modéliser l'application de détection d'intrusion en utilisant les concepts de détection d'intrusion de façon plus intuitive.

2.7. Conclusion

Dans ce chapitre nous avons présenté l'état de l'art de la détection anomalies et son application dans le domaine de la détection des attaques et des intrusions réseau. En fait, plusieurs types d'anomalies réseau résultent des problèmes de sécurité tels que les attaques et les intrusions. Pour faire face contre ces problèmes, différentes techniques et systèmes ont été mis en place, en se basant sur différentes familles de méthodes, telles que les méthodes statistiques, techniques de classification, le clustering et outliers, le soft computing et à base de connaissance. Pour évaluer leur qualité, diverses métriques sont utilisées, à savoir accuracy, FPR, TPR, TNR et FNR. Par ailleurs, une meilleure technique ou un système de détection devraient être en mesure de reconnaître les attaques et les intrusions dans les plus brefs délais que possible, afin que la réponse de défense puisse identifier les origines des

anomalies et d'initier le processus de défense approprié. En plus, ils doivent être capables d'identifier à la fois les attaques connues et inconnues avec de faibles taux de fausses alertes.

Malgré ces efforts considérables des acteurs du domaine et le nombre important des techniques et des systèmes ainsi développés, il existe encore un nombre de questions et de défis de recherches ouverts tels que le problème des fausses alarmes, l'établissement des seuils de détection et la dépendance de larges hypothèses des caractéristiques du trafic réseau. Dans le chapitre trois, nous investiguons l'utilité des cartes de contrôle dans la détection des attaques DOS et DDOS. Une étude comparative entre les cartes de contrôle conventionnelles sera présentée.

Chapitre3

Détection d'anomalies via les cartes de contrôle : application à la détection des cybers-attaques DOS et DDOS dans les réseaux IP

Dans ce chapitre

- Les cartes de contrôle
- Les différents types des cartes de contrôle
- La base de trafic DARPA99
- Détection des attaques DOS et DDOS par les cartes de contrôle
- Résultats de détection et discussions

Chapitre 3

Détection d’anomalies via les cartes de contrôle : application à la détection des cybers-attaques DOS et DDOS dans les réseaux IP

Récemment, le contrôle et l'amélioration de la qualité des produits et des services ont été considérés comme l'une des stratégies commerciales les plus importantes pour atteindre la satisfaction des clients à l'échelle mondiale. Dans ce sens, le contrôle statistique des processus (SPC : Statistical Process Control) offre une collection d'outils statistiques et analytiques qui peuvent être utilisés pour atteindre la stabilité du processus et réduire la variabilité de sa valeur cible. Ces outils jouent un rôle crucial pour détecter si un processus est sous contrôle (in-control) ou hors contrôle (out-of-control) [82].

L'idée de base du SPC est qu'un processus reste toujours dans un état de contrôle statistique (in-control), sauf si un événement spécial se produit. Un état de contrôle statistique existe lorsque certaines variables critiques du processus restent dans le voisiage de leurs valeurs cibles et ne changent pas de façon perceptible. La seule variation devrait être des fluctuations quotidiennes marginales. Autrement dit, un processus est sous contrôle statistique lorsque toutes les causes spéciales (assignable causes) de variation ont été éliminées et que seule la variation de cause naturelle (common causes) subsiste. Les cartes de contrôles sont des techniques de surveillance (monitoring) en ligne des processus, largement utilisée à cette fin [83] .

Les cartes de contrôle constituent un outil principal utilisé pour SPC, elles permettent une représentation graphique de certaines caractéristiques statistiques pour des mesures quantitatives spécifiques du processus surveillé.

L'un des principaux objectifs d'une carte de contrôle est de « détecter rapidement l'occurrence de causes attribuables (assignable causes) aux changements de processus afin que le processus puisse être étudié et que des mesures correctives soient prises avant la fabrication de nombreuses unités non conformes» [83].

Pendant de nombreuses décennies, l'utilisation principale des cartes de contrôle s'est concentrée sur les applications de contrôle de la qualité dans le domaine industriel. De nos jours, les cartes de contrôle ont été mises à bord de nombreux domaines, y compris les systèmes de santé [84] [85], l'économie [86], l'informatique [87], l'environnement [88] et la robotique [89].

D'autre part, la détection rapide et robuste des modifications et changements anormaux du trafic réseau est hyper importante pour de nombreuses fonctions de gestion du trafic (la planification et la gestion des allocations de bande passante, prévision des situations de congestion, la détection du trafic de déni de service…). Par conséquence, il peut être intéressant pour les administrateurs d'intervenir uniquement lorsqu'il se produit quelque chose d'inhabituel (c'est-à-dire une cause attribuable) ou, alternativement, une menace majeure est claire (par exemple, une attaque en cours).

L'objectif de ce chapitre est de montrer l'utilité des cartes de contrôle dans la détection des cyber-attaques y compris les attaques DOS et DDOS. Précisément, nous introduisons ici une étude comparative entre les cartes de contrôle les plus utilisées (Shewhart, CUSUM et EWMA) lors de la détection des attaques TCP SYN flood, UDP flood et les attaques SMURF. La comparaison est menée en utilisant les traces de trafic IP de la base DARPA99 (Defense Advanced Research Projects Agency 99).

Après l'introduction du concept général et le principe de fonctionnement des cartes de contrôles, nous dériverons, en particulier, les cartes les plus répandues, à savoir : Shewhart, CUSUM et EWMA. La base de trafic DARPA99 et notre méthode de prétraitement pour extraire les caractéristiques du trafic nécessaires à la détection des différents types d'attaques seront ainsi présentées. Ensuite, nous introduirons la procédure adoptée pour la détection des attaques DOS et DDOS à base d'anomalies via les cartes de contrôle. En fin, nous les validerons sous différents scénarios d'attaques, en analysant et comparant leurs performances de détection.

3.1. Définition et principe de fonctionnement des cartes de contrôle

Les cartes de contrôle remontent à Walter Shewhart, des Bell Telephone Laboratories, qui les a introduit pour la première fois en 1924 et les a utilisé pour instaurer les bases du SPC modern [83] [90].

L'hypothèse de base sous-jacente aux cartes de contrôle est que la variation de la qualité des produits et des services est due en partie à des causes communes courantes (common causes) et en partie à des causes spéciales ou causes attribuables (assignable causes) [83] [91]. Le terme cause commune fait référence à la variabilité inhérente ou naturelle qui est présente dans un processus. Il s'agit de légers changements (on parle aussi de bruit de fond) dans les caractéristiques statistiques du processus, incontrôlables ou toujours présent et qui pourraient être dues à l'effet cumulatif de nombreuses causes petites et indétectables (mais inévitables). Les causes spéciales de variation sont dues à un incident survenu dans le processus et qui découle d'une ou d'ensemble de sources : matériels, logiciels, humaines et environnementales. Cette variabilité est généralement importante par rapport à la variabilité inhérente. Elle affecte directement la qualité du processus et représente généralement un niveau inacceptable de performances.

En fait, l'objectif principal d'une carte de contrôle est de séparer les causes communes des causes spéciales. Lorsqu'une cause particulière de variation est détectée, le processus est arrêté (ou autre intervention adéquate est installée) et des investigations sont menées pour déterminer la source et l'éliminer (si possible).

Etant donné que des causes assignables entraînent des changements dans les paramètres du processus, les cartes de contrôle sont utilisées pour détecter tout changement dans un ou plusieurs de ces paramètres. La carte de contrôle est construite en prélevant des échantillons (ou des mesures) avec une cadence donnée et tracer ensuite la statistique appropriée pour chaque échantillon (moyenne, écart-type, étendue, nombre, pourcentage), comme l'illustre la figure 3.1. Elle comprend trois lignes principales : la ligne centrale (CL : Central Line), la limite de contrôle inférieure (LCL : Lower Control Limit) et la limite de contrôle supérieure (UCL : Upper Control Limit). Tant que les valeurs de la statistique de l'échantillon se situent entre les deux limites LCL et UCL, le processus est considéré sous contrôle. Cependant, si les points des échantillons se placent en dehors de ces limites, le processus est alors hors de contrôle. Les valeurs d'UCL et LCL sont généralement choisies de telles sortes que lorsque le processus est sous contrôle, la probabilité qu'un point soit tracé en dehors de ces limites est très faible [83].

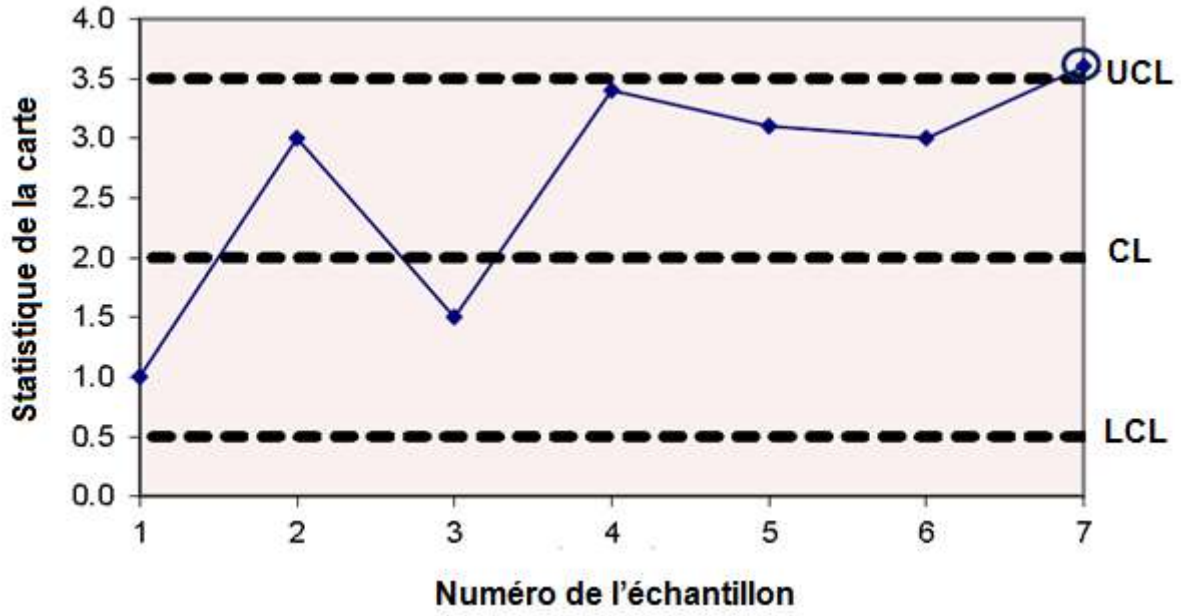

Figure 3.1 : Exemple d'une carte de contrôle

3.2. Les différents types de cartes de contrôle

En se basant sur le type de la caractéristique statistique tracée, on peut distinguer trois types de cartes de contrôle : Shewhart, CUSUM et EWMA. Les cartes de contrôle EWMA et CUSUM sont des cartes avec mémoire (memory-based) qui combinent plusieurs échantillons passés avec l'échantillon présent (ou actuel) dans le processus de décision. La taille du mémoire (courte ou longue) dépend des paramètres d'ajustement qui caractérisent chaque carte. La carte Shewhart, cependant, utilise uniquement les informations disponibles à partir de l'échantillon le plus récent, le dernier.

3.2.1. Les cartes Shewhart

Les cartes de contrôle Shewhart sont bien connues comme les premières cartes de contrôle proposées dans l'histoire. Elles prennent le nom de leur inventeur Walter A. Shewhart, qui les a développé en 1924 pour détecter les changements d'une variable aléatoire [83]. Elles constituent le type de cartes le plus populaire dans la pratique, en raisons de leur simplicité, leur facilité d'implémentation et du fait qu'elles sont assez efficaces pour détecter les déviations modérées à importantes [92].

La carte Shewhart permet une représentation graphique de base des valeurs successives d'une caractéristique statistique calculées à partir des échantillons de mesures,

prises sur le paramètre à surveiller, et tracées en fonction du nombre ou du temps de prélèvement de l'échantillon. La figure 3.2 montre la conception d'une carte de contrôle Shewhart pour la surveillance d'un paramètre de processus. L'axe horizontal représente le numéro ou le temps de prélèvement de l'échantillon. L'axe vertical représente la valeur de la caractéristique statistique calculée pour chaque l'échantillon.

Pour des échantillons Xi (i=1… n) indépendants et identiquement distribués, d'un processus normal $\mathcal{N}(\mu, \sigma^2)$, Shewhart définit les limites de contrôle suivantes [83] :

$$CL = \mu_0 \qquad (3.1)$$

$$UCL = \mu_0 + 3\sigma \qquad (3.2)$$

$$LCL = \mu_0 - 3\sigma \qquad (3.3)$$

Avec μ_0 , σ sont la moyenne et l'écart type dans le cas où le système surveillé est sous-contrôle.

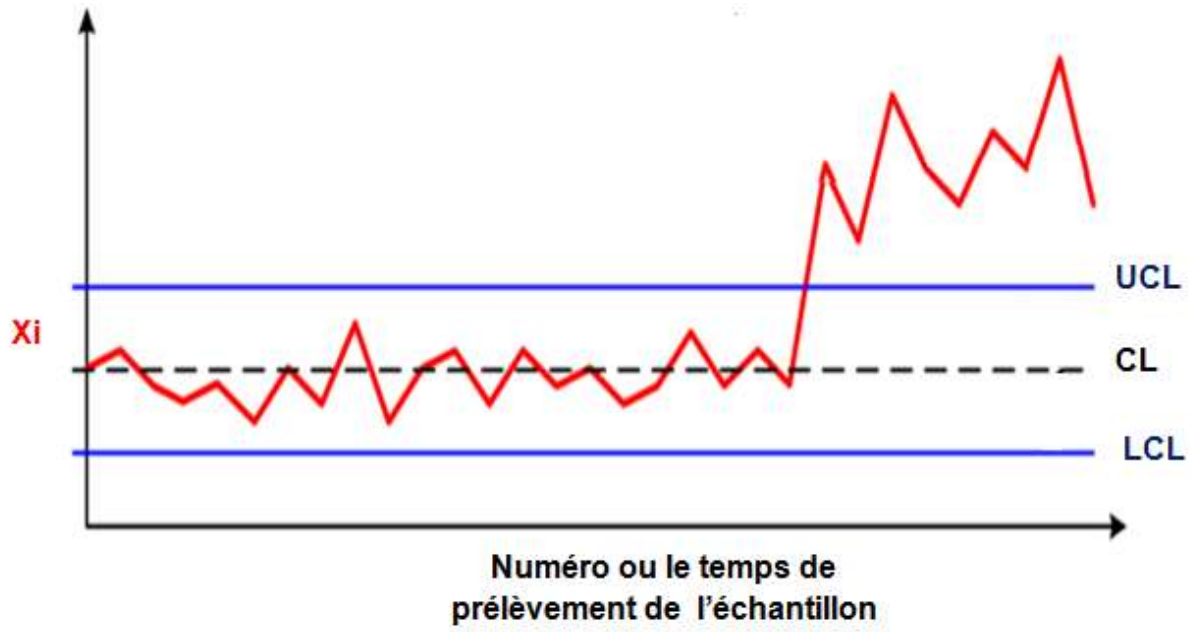

Figure 3.2 : Principe de la carte de contrôle Shewhart

3.2.2. Les cartes CUSUM

Les cartes de contrôle CUSUM ont été introduites par E.S. Page en 1961 [93]. Elles sont préférées pour identifier les causes persistantes dans le processus surveillé. L'idée de base du CUSUM est que les petits changements deviennent plus détectables si on accumule leurs écarts par rapport à la valeur cible de la statistique considérée. En incorporant directement tous les échantillons de mesures prélevés (carte avec mémoire), CUSUM consiste à

représenter la somme cumulée de la différence entre les observations (les valeurs des échantillons) Xi et la valeur moyenne cible. Lorsque le processus est sous contrôle, μ_0 peut être considéré comme étant la valeur cible pour les caractéristiques de Xi. La somme cumulative C_i jusqu'au $i^{ième}$ échantillon est donnée par [83] :

$$C_i = \sum_{j=1}^{i}(x_j - \mu_0) \qquad (3.4)$$

Dans la version tabulaire, CUSUM définit deux accumulations unilatérales : positive C_i^+ et négative C_i^- [83]:

$$C_i^+ = \max [0, Xi - (\mu_0 + k) + C_{i-1}^+] \quad (3.5)$$

$$C_i^- = \max [0, (\mu_0 - k) - Xi + C_{i-1}^-] \quad (3.6)$$

C_i^+: représente l'accumulation des dérivations au-dessus de la moyenne cible.

C_i^-: c'est l'accumulation des dérivations au-dessous de la moyenne cible.

$C_i^+ = C_i^- = 0$ sont les valeurs initiales**.**

k: un paramètre d'ajustement de la sensibilité de CUSUM, choisi de façon à restreindre le nombre de fausses alertes. Il dépend de l'importance de l'écart que l'on souhaite révéler. Plus k est petit, plus la carte sera capable de détecter les faibles changements mais plus on augmente le risque de fausses alertes. Il souvent calculé par :

$$k = \frac{\sigma}{2} \quad (3.7)$$

Les limites de contrôle UCL et LCL utilisées par la carte CUSUM sont définies par l'intervalle de décision H :

$$H = 5 \times \sigma \quad (3.8)$$

Lorsque l'un des deux entités C_i^+ ou C_i^- dépasse H, le processus est considéré hors contrôle. CUSUM est, particulièrement adaptée pour la détection de faibles changements dans un processus.

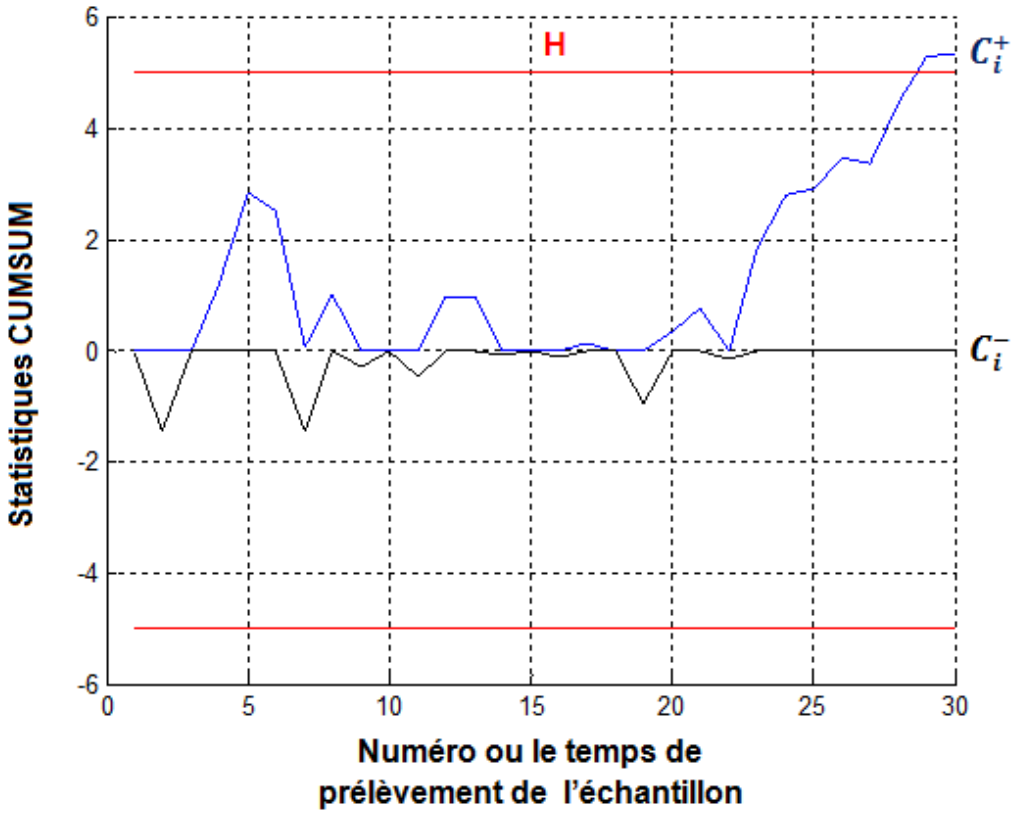

Figure 3.3 : Principe de la carte CUSUM (version tabulaire)

3.2.3. Les cartes EWMA

Les cartes EWMA ont été introduites par S.W. Roberts (1959) [94] et constituent également une bonne alternative de la carte de contrôle Shewhart pour détecter les petits changements. Etant donné que l'EWMA peut être considérée comme une moyenne pondérée de toutes les observations antérieures et actuelles, EWMA est un algorithme dynamique qui adapte constamment sa valeur en fonction des mesures reçues. Elle dépend d'un facteur de poids de lissage qui détermine avec quelle vitesse les anciennes valeurs (ou les plus âgées) seront négligées.

La carte EWMA accumule, avec une pondération exponentielle, les échantillons prélevés, en donnant les poids élevés aux échantillons plus récents. EWMA est définie par la formule suivante [83]:

$$Z_i = \lambda x_i + (1 - \lambda) Z_{i-1} \quad (3.9)$$

Pour montrer que Z_i est une moyenne pondérée, (3.9) peut être reformulée comme suit :

$$Z_i = \lambda \sum_{j=0}^{i-1} (1-\lambda)^j x_{i-j} + (1-\lambda)^i Z_0 \quad (3.10)$$

Les limites de contrôles pour EWMA sont calculées par [83]:

$$UCL = \mu_0 + L\sigma \sqrt{\left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2i}]} \quad (3.11)$$

$$CL = \mu_0 \quad (3.12)$$

$$LCL = \mu_0 - L\sigma \sqrt{\left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2i}]} \quad (3.13)$$

Avec:

x_i: Valeur d'i^{ieme} échantillon

Z_{i-1} : EWMA de l'échantillon précédent

Z_0: valeur initial de l'EWMA, généralement, choisie égale à μ_0 .

λ $(0 < \lambda \leq 1)$: facteur d'ajustement compris entre 0 et 1. Il représente le poids attribué aux different échantillants [83]:

- **Plus λ est proche de 0 :** le poids $\lambda(1-\lambda)^j$ diminue lentement, et plus on tient compte du passé. Cela implique que l'on identifiera plus facilement les faibles dérives. Par contre, les dérives brutales et les déréglages importants, seront moins bien détectés.
- **Plus λ est proche de 1 :** $\lambda(1-\lambda)^j$ diminue rapidement, et moins on tient compte du passé. Cela implique que l'on aura une meilleure réactivité pour identifier les déréglages brusques mais à contrario, on détectera moins bien les faibles variations.
- **Si λ = 1 :** EWMA sera équivalente à la carte de contrôle classique Shewhart.

L : La largeur des limites de contrôle.

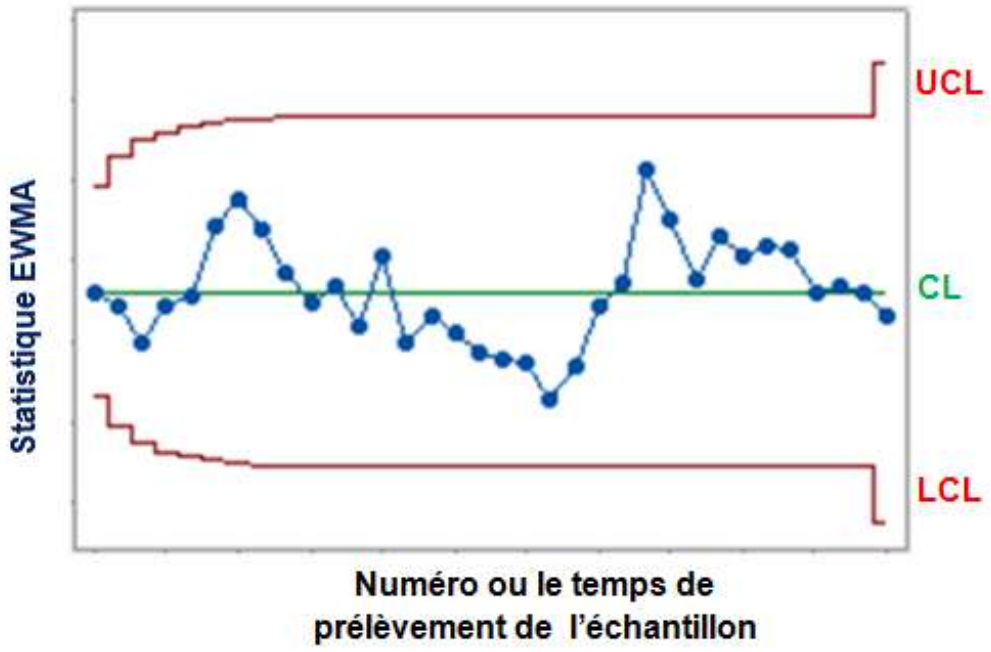

Figure 3.4 : Principe de la carte EWMA

3.3. Détection des attaques DOS et DDOS par les cartes de contrôle

Les caractéristiques ou les paramètres du trafic réseau (nombre messages, nombre segments, nombre de bits, protocoles, les ports, les adresses IP…etc.) peuvent subir des changements significatives lors des attaques DOS et DDOS. Les cartes de contrôle peuvent être utilisées pour révéler leurs déviations par rapport à l'état normal et permettent par conséquence d'identifier les trafics anormaux résultant des différents types des attaques DOS et DDOS.

Ainsi, pour utiliser les cartes Shewhart, CUSUM et EWMA pour détecter les attaques DOS et DDOS, la procédure générale sera :

① **Capture et collection du trafic réseau**

Le trafic réseau contient des paquets de données et de contrôle acheminés sur les liaisons de communication entre les différents équipements du réseau et entre les machines terminaux des utilisateurs, reflétant ainsi les activités sur les réseaux exploités. Plusieurs outils de capture du trafic peuvent être utilisés comme le TCPDUMP [95], Wireshark [97], SolarWinds Deep Packet Inspection and Analysis tool [97],...Ces outils peuvent être installés à travers le réseau à protéger, et permettent la récupération online du trafic qui les traverse. Le trafic capturé peut être ensuite stocké pour des futures utilisations, y compris la détection des différente formes d'anomalies, d'intrusions et les éventuelles attaques DOS et DDOS.

② Prétraitement et extraction des paramètres

Etant donné que les attaques DOS et DDOS peuvent affecter un ou plusieurs paramètres du trafic réseau, une étape cruciale dans les techniques de détection à base d'anomalie est la sélection des paramètres à contrôler. Cependant, le trafic réseau est généralement capturé sous sa forme brute, avec la totalité des échanges (différents protocoles, différents messages,…), d'où un prétraitement est nécessaire. En fait, cette étape permet, d'une part, l'extraction et l'isolation des paramètres de détection, et d'autre part de les préparer et les transformer en données (mesurables) d'entrée utilisable dans les cartes de contrôle (cf paragraphe 3.4.2*).*

③ Construction de la carte de contrôle

Une fois les données et les paramètres du trafic sont préparés, les cartes de contrôle peuvent être construites pour la détection des différents types des attaques DOS et DDOS. Les données d'apprentissage issu du trafic normal en termes de paramètres de détection sont utilisés pour calculer les limites de contrôle CL/UCL/LCL avec les trois cartes de contrôle : Shewhart (équations 3.1, 3.2, 3.3), CUSUM (équation 3.8) et EWMA (équation 3.11, 3.12, 3.13). Ces limites definissent donc l'intervalle de variation des paramètres surveillés dans le fonctionnement normal du réseau, et dans l'absences des attaques ciblées.

④ Détection et identification des attaques

Des données de test qui peuvent contenir des attaques DOS et DDOS, on calcul la statistique caractéristique de chaque carte, c à d, la valeur de l'échantillon, C_i, C_i^+ et C_i^- et Z_i pour Shewhart, CUSUM et EWMA, respectivement. Ces caractéristiques sont ensuite, comparées avec les limites de contrôle :

⇒ Si ces caractéristiques sont comprises entre UCL et LCL, alors le trafic surveillé est considéré comme normal.

⇒ Sinon, si ces caractéristiques dépassent l'une des limites de contrôle (UCL ou LCL), le trafic, dans ce cas, a un comportement anormal, une alarme de détection d'attaques DOS ou DDOS est déclenchée, en donnant ainsi les informations liées aux attaques détectées (type, victime et la date d'apparition).

La figure 3.5 récapitule la procédure générale d'utilisation des cartes de contrôle Shewhart, CUSUM et EWMA pour la détection des attaques DOS et DDOS.

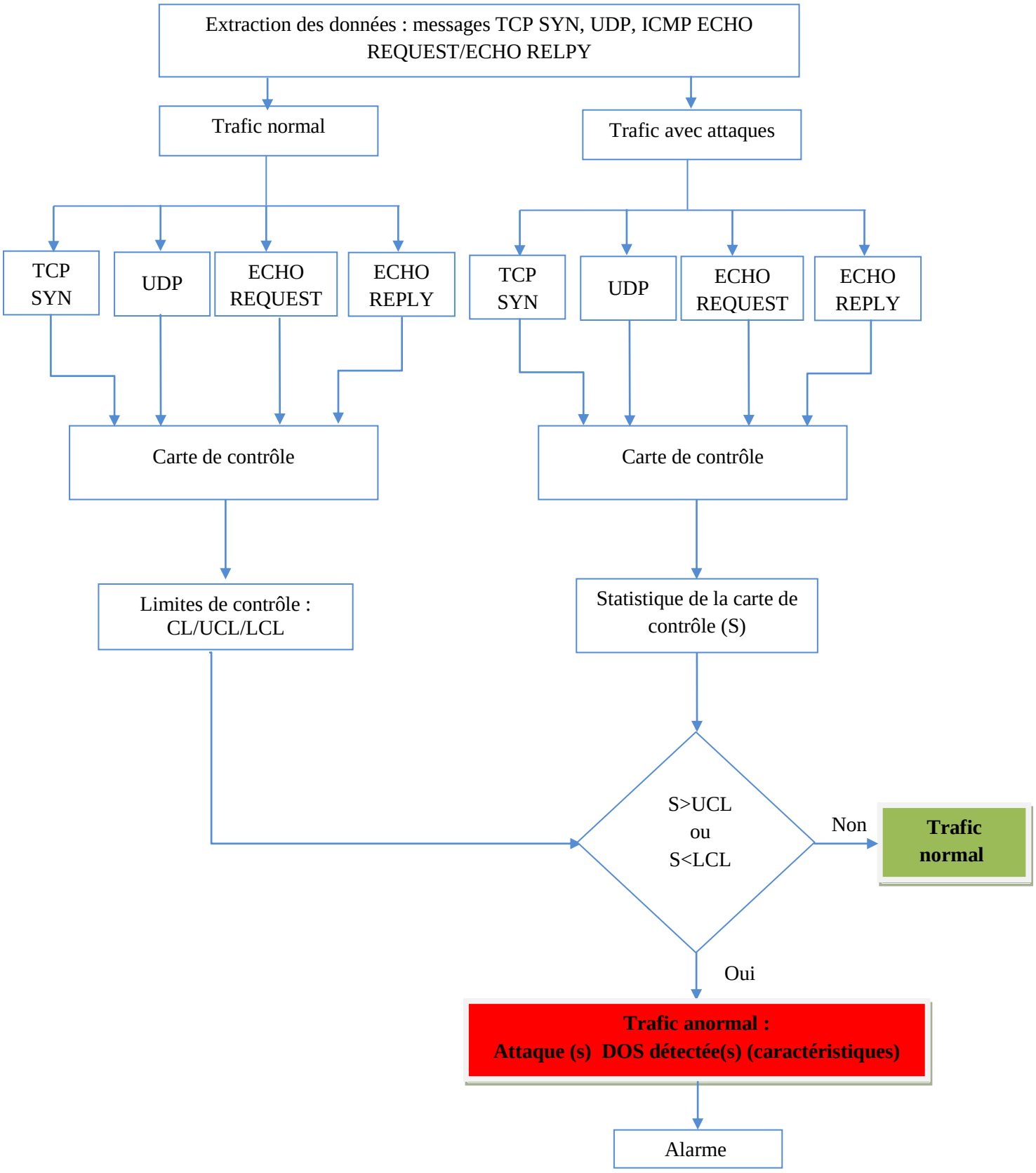

Figure 3.5 : Procédure générale de détection des attaques DOS et DDOS par les cartes de contrôle Shewhart, CUSUM et EWMA

3.4. La base de trafic DARPA99

3.4.1. Présentation générale de la base DARPA99 [98]

Pour évaluer les performances des cartes Shewhart, CUSUM et EWMA, nous utiliserons la base de données DARPA99. DARPA 99 (appelé aussi IDEVAL : Intrusion Detection Evaluation) s'inscrit parmi les bases de données les plus importantes et les plus utilisées pour l'évaluation des systèmes de détection d'intrusion. Elle a été construite par Lincoln Laboratory du Massachusetts Institute of Technology (MIT) sous le support de la DARPA et de l'Air Force Research Laboratory (AFRL). Elle représente le trafic capturé dans un réseau réel similaire au réseau réel d'une base militaire de l'Air Force, connectée à Internet.

La figure 3.6 montre la topologie du réseau mis en place. Les machines à gauche simulent le réseau de la base militaire (le réseau intérieur), tandis que les machines à droite simulent l'Internet (le réseau extérieur). Un routeur Cisco connecte les deux réseaux.

Les victimes sont des serveurs de la base de l'Air Force qui font les cibles des différents types d'attaques. Pascal exécutant Solaris 2.5, un shell ou un serveur de connexion fournit des services telnet, SMTP, SSH et FTP. Zeno, fonctionne avec SunOS 4.1.4, est un serveur de fichiers, sendmail et permettait le partage de fichiers entre les utilisateurs via un serveur FTP. Marx, travaille avec RedHat 5.0, est le serveur Web, il sert comme page d'accueil à la fois pour Internet et pour une utilisation interne, un serveur Web Apache a été utilisé. Hume, c'est un serveur Windows NT 4.0, était équipé avec IIS (Internet Information Server) et héberge des serveurs FTP, gopher, Web et plusieurs autres utilitaires incluent un serveur de messagerie, appelé MailSrv. Kant utilise un Windows98. Le serveur Web, Aesop, exécutant RedHat 5.0 est le serveur Web Internet et semble être des milliers de serveurs Web Internet individuels. Les hôtes virtuels internes et externes sont utilisés pour usurper différentes adresses IP.

Deux postes de travail, l'un avec Linux RedHat 5.0 et l'autre avec Windows NT4.0, sont utilisés comme attaquants internes. Trois autres postes de travail, deux avec Linux RedHat 5.2 et un avec Windows NT 4.0, sont utilisés comme attaquants extérieurs. Plusieurs types d'attaques générées dans ces données, y compris les attaques de déni de service.

Pour collecter le trafic réseau, deux renifleurs ont été installés sur le réseau. Locke, le renifleur interne, exécute Solaris 2.6 et utilisé pour capturer le trafic réseau sur le réseau intérieur. Le renifleur extérieur, Solomon, exécute également Solaris 2.6 et est utilisé pour capturer le trafic réseau (entrant/sortant) vers et depuis la base de l'Air Force. Pour les deux

renifleurs, UNIX TCPDUMP a été utilisé pour collecter le trafic. Cinq semaines de données ont été collectées. Chaque semaine comprenait cinq jours, du lundi au vendredi, avec 22 heures par jour, de 8h00 à 6h00.

Plus de 8Goctets de trafic réseau (entrant et sortant) sous forme de fichiers compressés TCPDUMP a été enregistré. Pour faciliter l'évaluation des IDS à base d'anomalies, le trafic collecté contient trois semaines de données d'apprentissage (training data) séparées de deux semaines de données de test. Les première et troisième semaines de données d'apprentissage sont totalement exemptes d'attaques, la seconde en inclue certaines. Les deux semaines de données de test (semaines quatre et cinq) contiennent différentes catégories d'attaques.

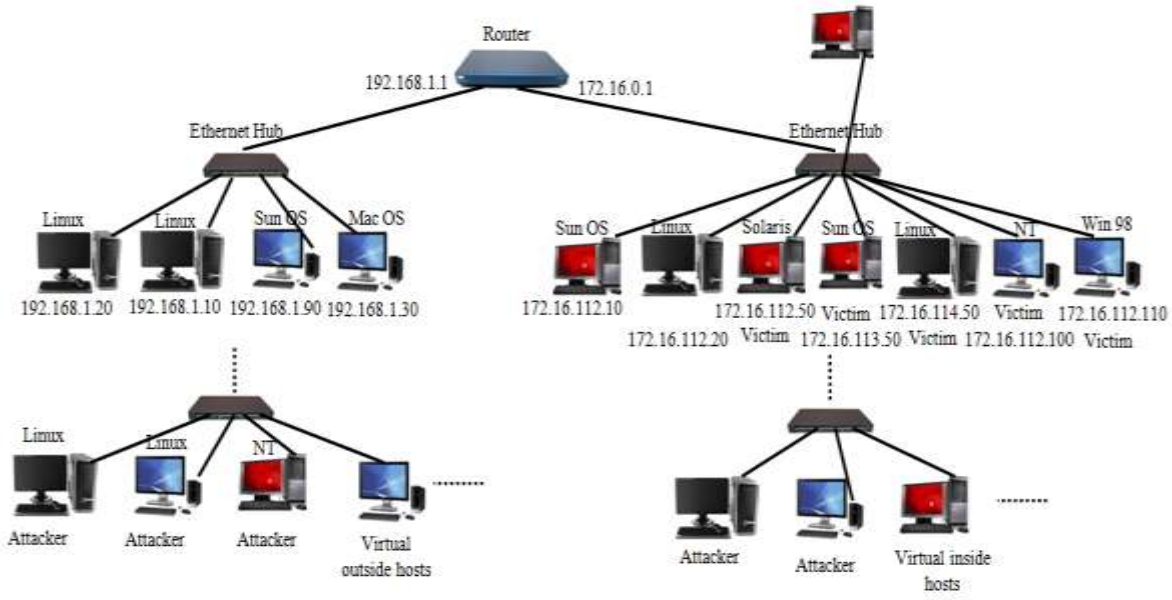

Figure 3.6 : La topologie de réseau utilisé par DARPA 99 [98]

3.4.2. Prétraitement et extraction des paramètres de détection

Etant donnée que la base DARPA99 fournit le trafic total capturé dans le réseau (trames, paquets, différents protocoles, différents messages,…) (figure 3.7), un prétraitement de la base est nécessaire pour extraire les paramètres de détection et les préparer pour qu'on puisse les utiliser aux entrées des cartes de contrôle. Vu que les attaques DOS et DDOS de type TCP SYN flood, UDP flood et SMURF sont basées sur l'envoi massif des segments SYN, des datagrammes UDP, des messages ICMP ECHO-REPLY respectivement, le prétraitement des traces de trafic consiste en l'extraction, le traitement et la préparation de ces différentes structures de données.

A partir des fichiers TCPDUMP de DARPA99, nous avons utilisé l'analyseur réseau Wireshark pour filtrer les paramètres en question (segments SYN, datagrammes UDP, messages ICMP ECHO REPLY). Ce filtrage nous a permis d'obtenir des fichiers qui contiennent seulement ces paramètres. La figure 3.8 donne l'exemple de filtrage des segments TCP SYN. Ensuite nous avons utilisé MySql et Java pour traiter les fichiers résultants de Wireshark. Ce traitement permet d'une part d'effectuer de nombreuses opérations (nombre de messages par unité de temps ou par intervalle de mesure, par destination,…) et d'autre part de les transformer en données de formats qui peuvent être directement manipulé par Matlab ou les autres outils de calcul. La figure 3.9 illustre la forme finale, après le prétraitement, des segments TCP SYN et des messages ICMP ECHO-REPLY.

Cette procédure de prétraitement nous a permis, ainsi, de construire deux nouvelles bases de données, de quelques Ko, sous formats .cvs et .txt :

- La première : contient le trafic DARPA99 en termes de segments SYN, des datagrammes UDP et des messages ICMP ECHO-RPLY, seulement.
- La deuxième : contient un trafic synthétisé, inspiré de la première base, et constituée de trois types de trafics : trafic normal, trafic avec attaques TCP SYN flood, UDP flood et SMURF de différentes intensités (faible et élevée).

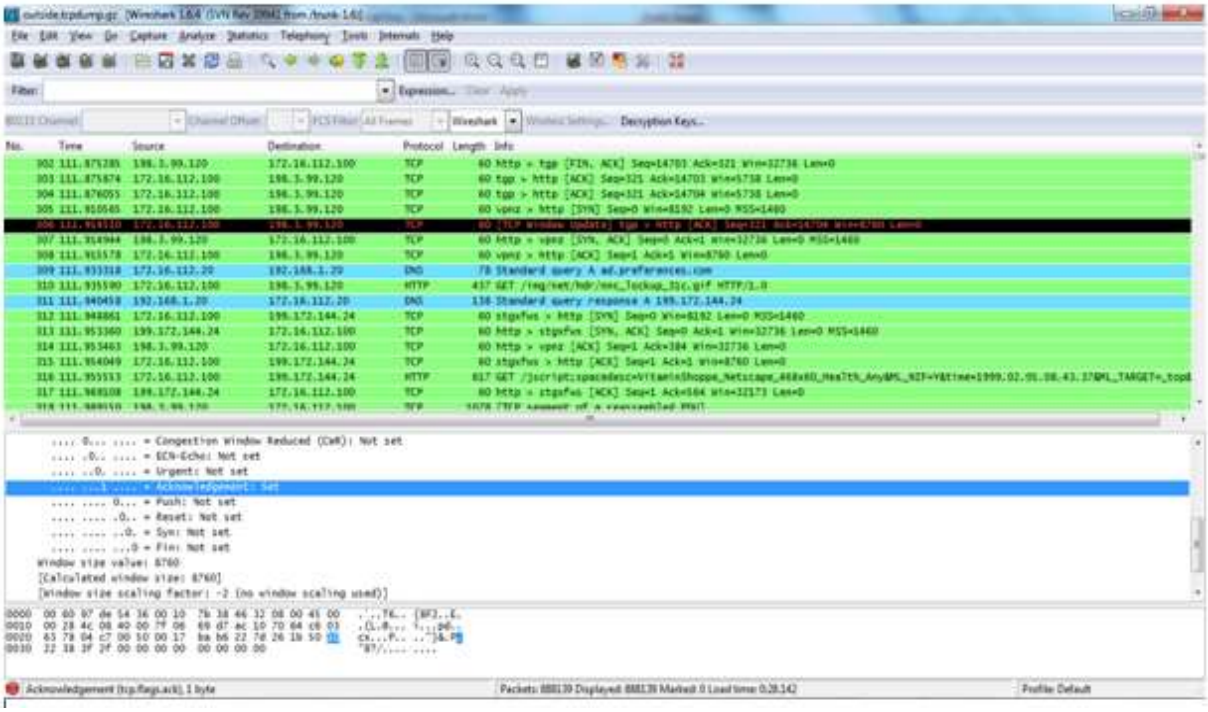

Figure 3.7 : Trafic brute DARPA99 visualisé avec Wireshark

(exemple : semaine2/jour3)

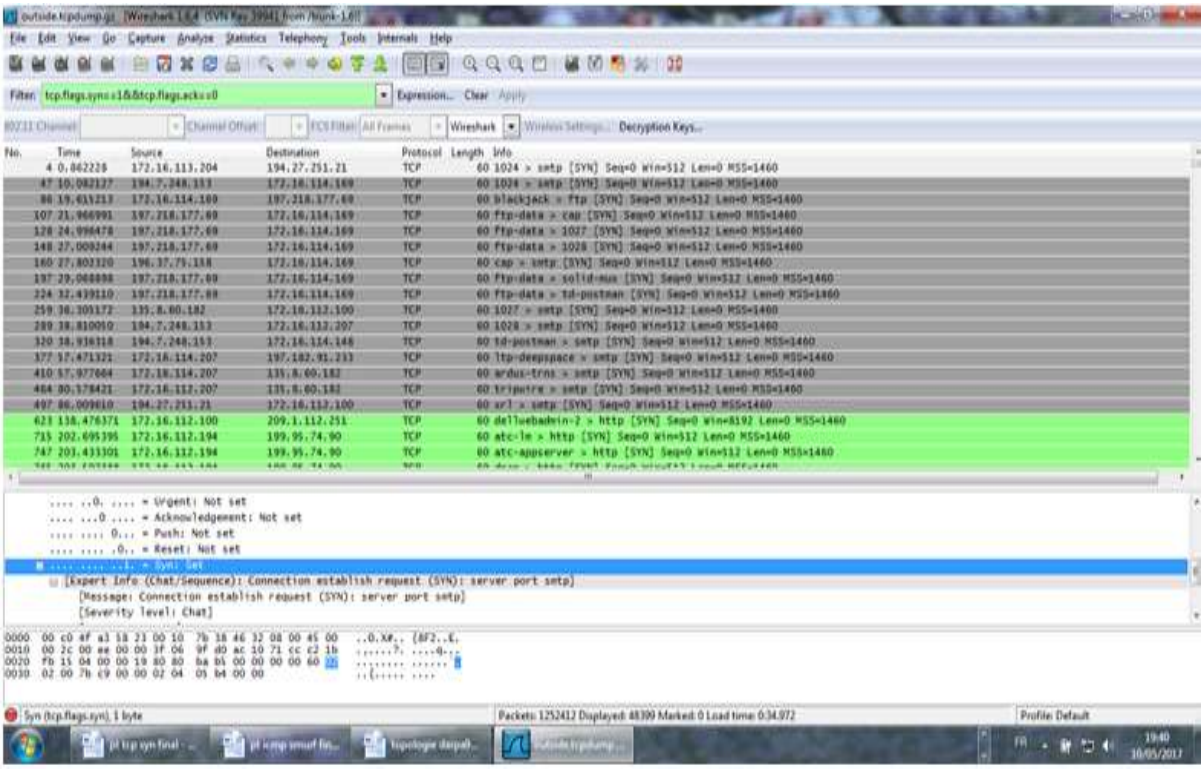

Figure 3.8 : Filtrage des segments SYN avec Wireshark

N23Rauto - Bloc-notes

Fichier Edition Format Affichage ?

timet	Syn
0	0
10	0
20	0
30	0
40	0
50	0
60	5
70	0
80	0
90	2
100	0
110	0
120	12
130	0
140	0
150	11
160	0
170	35
180	6
190	0
200	0
210	0
220	26
230	0
240	0
250	1
260	0
270	0
280	0
290	0
300	0
310	0
320	1
330	0
340	0

a)

icmpn25rinreplyauto - Bloc-notes

Fichier Edition Format Affichage ?

time	echo
0	0
10	0
20	0
30	0
40	7
50	0
60	0
70	0
80	0
90	0
100	1
110	0
120	0
130	0
140	0
150	0
160	0
170	0
180	0
190	0
200	0
210	0
220	0
230	0
240	0
250	0
260	0
270	0
280	0
290	0
300	0
310	0
320	0
330	0
340	0

b)

Figure 3.9 : Exemples de paramètres de détection après prétraitement avec un intervalle de mesure de 10s a) les segments TCP SYN , b) les messages ICMP ECHO-REPLY

3.5. Résultats de détection

En se basant sur la méthode de détection des attaques DOS et DDOS par les cartes de contrôle, présentée dans le paragraphe 3.3, nous étudions l'efficacité des cartes Shewhart, CUSUM et EWMA dans la détection des attaques TCP SYN flood et SMURF. Une étude comparative entre les cartes est ainsi présentée. Nous avons effectué de nombreuses simulations, nous présentons ici une partie des résultats obtenus.

3.5.1. Détection des attaques TCP SYN flood

Dans cette étape, nous utilisons les trois cartes de contrôle pour la détection des attaques TCP SYN flood. Nous considérons trois types de trafics anormaux (de test): trafic avec des attaques TCP SYN de faible intensité, trafic avec attaques TCP SYN d'intensité élevée et le flux des segments SYN dans le deuxième jour de la semaine cinq du trafic DARPA99. Le trafic d'apprentissage est celui du jour 3 de la semaine 2. Les limites de contrôle (CL/UCL/LCL) établies avec les trois cartes sont reportées sur le tableau 3.1.

Trafic		Semaine 2, jour 3
Limites de contrôle UCL, LCL	Shewhart	UCL=322.8864 LCL=-316.7486
	CUSUM	-H=-533.0291 +H=-533.0291
	EWMA	UCL= 31.8525/41.7933/ 48.2701/ 52.9031/56.3614 59.0061/ 61.0599/62.6717/ 63.9459/64.9588 /65.7673/ 66.4146/ 66.9341/67.3518/ 69.1029/…/69.1030 LCL= -25.7146/ -35.6554/-42.1323/ -46.7652 / -50.2235/-52.8682 / -54.9221/ -56.5338/ -57.8081/ -58.8210/ -59.6294/ -60.2767/ -60.7962/ -61.2140/ -61.5503/ …/-62.9651

Tableau 3.1 : Les limites de contrôle UCL et LCL établies avec les cartes Shewhart, CUSUM et EWMA (données d'apprentissage : Semaine 2, jour 3)

Afin d'évaluer la capacité des cartes de contrôle à détecter les attaques de faible intensité, nous avons généré des attaques TCP SYN flood de faible intensité, toutes les 3 heures avec une durée de 2mn pour chaque attaque. La figure 3.10 représente l'évolution des segments SYN durant le trafic correspondant (trafic ASFI : Attaque SYN à Faible Intensité). Les figures 3.11, 3.12, 3.13 et 3.14 montrent les résultats de détection obtenus avec les cartes Shewhart, CUSUM et EWMA, respectivement.

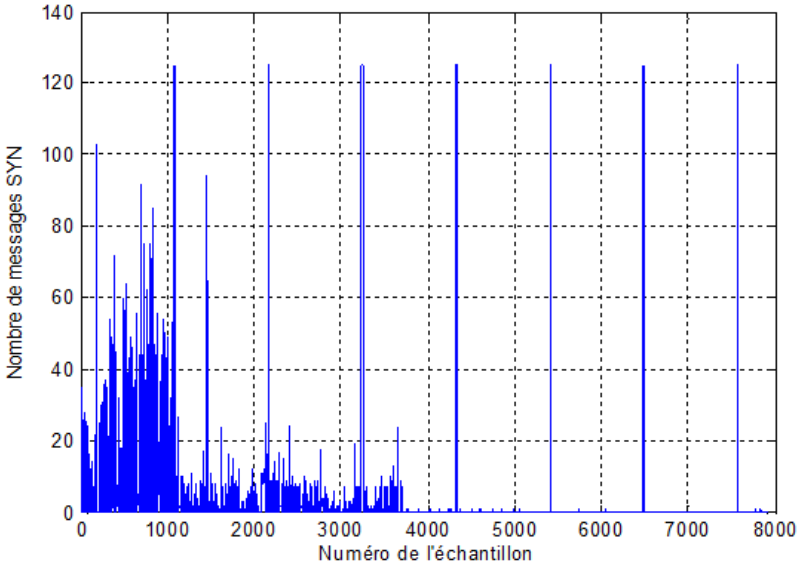

Figure 3.10 : Evolution du nombre de messages SYN en fonction du numéro de l'échantillon (Trafic ASFI)

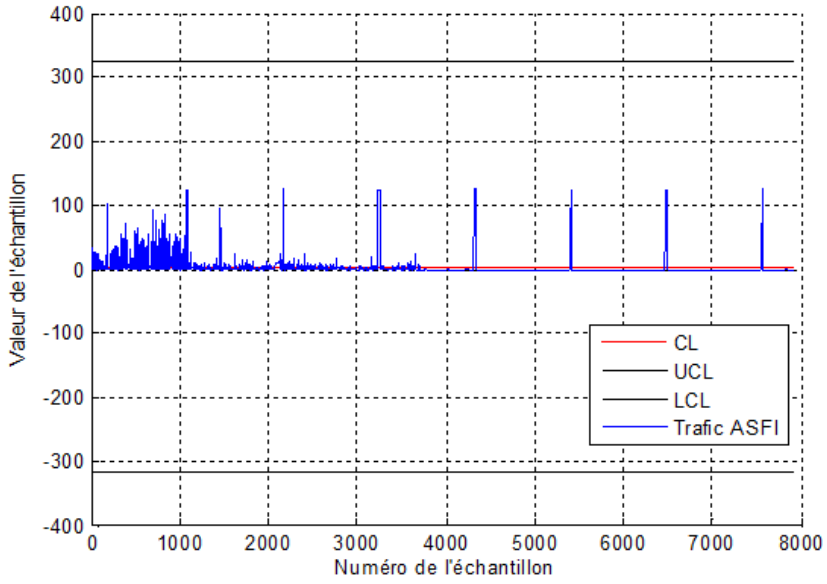

Figure 3.11 : Evolution des valeurs des échantillons en fonction du numéro de l'échantillon (Trafic ASFI)

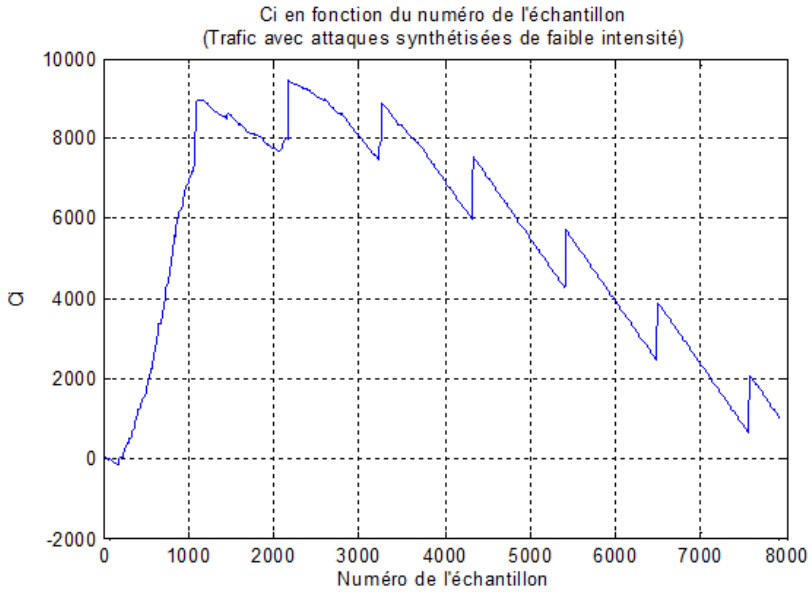

Figure 3.12 : Evolution du Ci en fonction du numéro de l'échantillon

(Trafic ASFI)

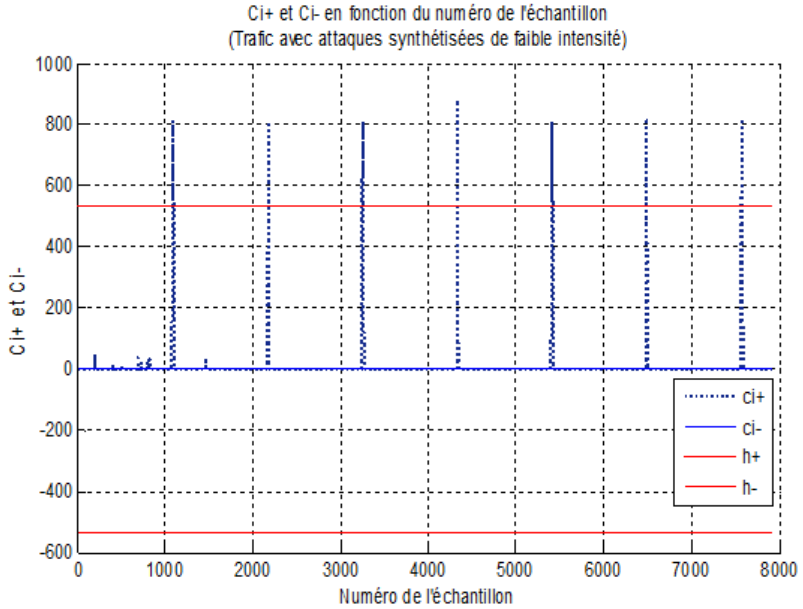

Figure 3.13 : Evolution des C_i^+et C_i^- en fonction du numéro de l'échantillon

(Trafic ASFI)

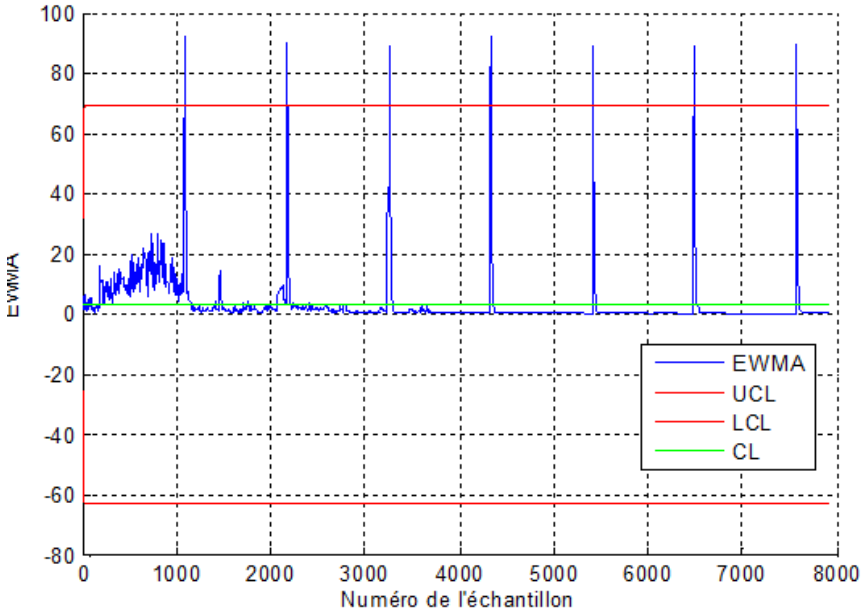

Figure 3.14 : Evolution de l'EWMA en fonction du nombre de l'échantillon (trafic ASFI)

Maintenant, les cartes seront testées avec un trafic qui contient des attaques TCP SYN d'intensité élevée (attaques de 2mn chaque 3 heures). La figure 3.15 illustre le trafic généré (trafic ASIE : Attaque TCP SYN d'Intensité Elevée), et les figures 3.16, 3.17, 3.18 et 3.19 récapitulent les résultats de détection obtenus avec les trois cartes.

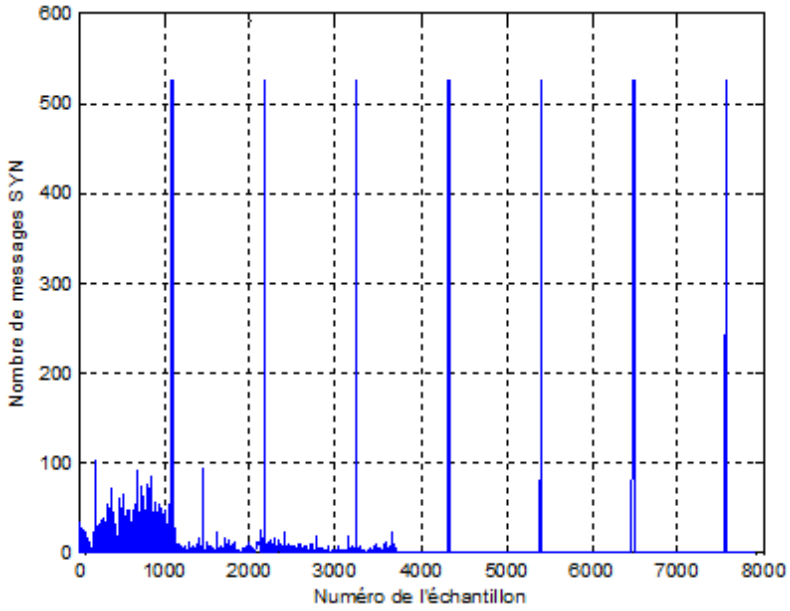

Figure 3.15 : Evolution du nombre de messages SYN en fonction du numéro de l'échantillon (Trafic ASIE)

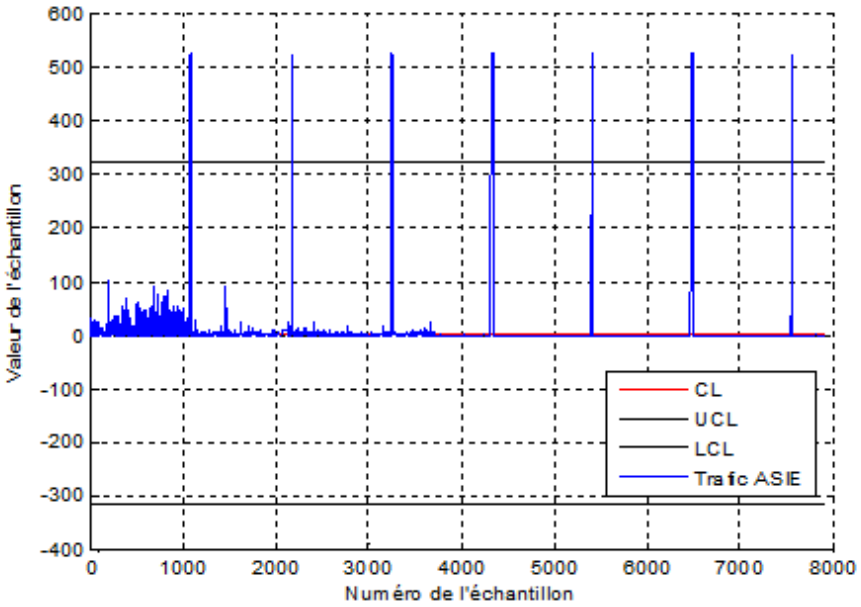

Figure 3.16 : Evolution des valeurs des échantillons en fonction du numéro de l'échantillon (Trafic ASIE)

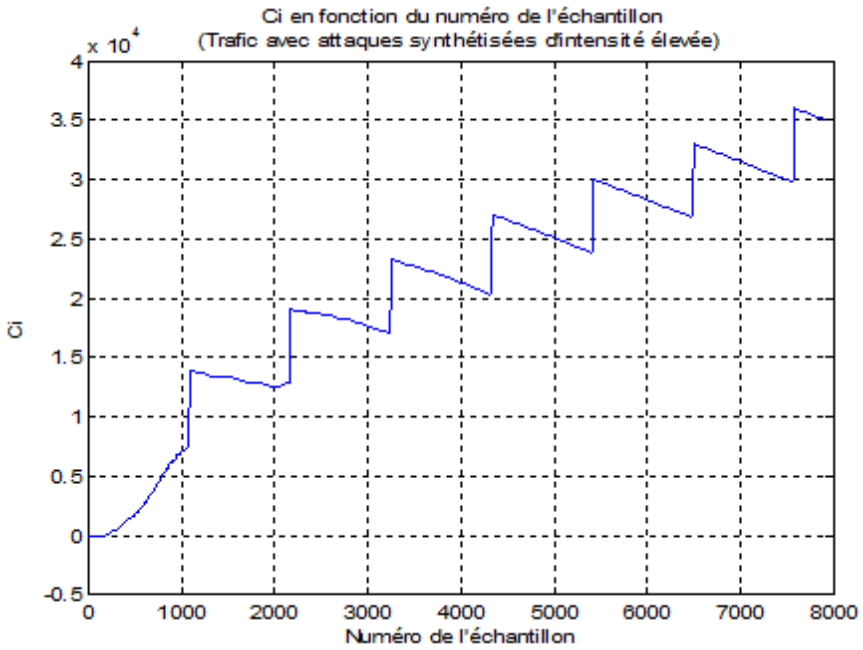

Figure 3.17 : Evolution du Ci en fonction du numéro de l'échantillon (Trafic ASIE)

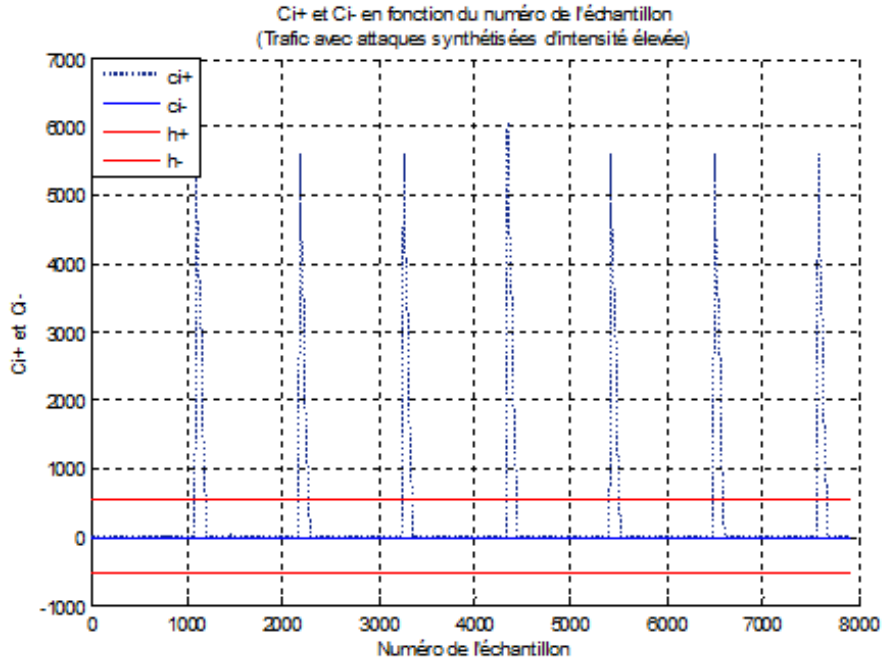

Figure 3.18 : Evolution des C_i^+ et C_i^- en fonction du numéro de l'échantillon (Trafic ASIE)

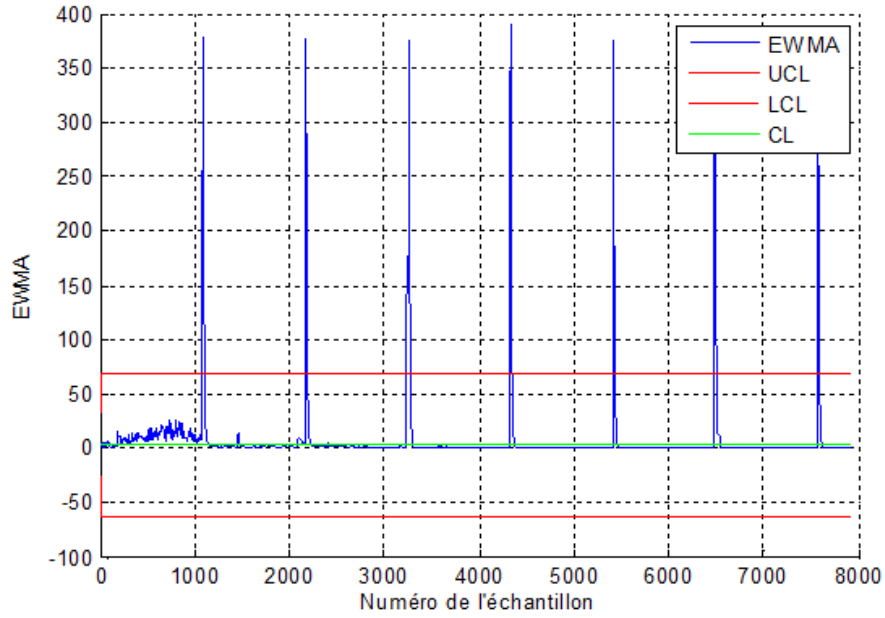

Figure 3.19 : Evolution de l'EWMA en fonction du numéro de l'échantillon (trafic ASIE)

Finalement, on applique les trois cartes aux données de test DARPA99 qui contiennent deux attaques TCP SYN flood dans le jour 2 de la semaine 5, dont les durée sont respectivement, 13mn 41s et 3mn 26s. Les figures suivantes présentent ce trafic de test ainsi que les résultats de détection correspondants.

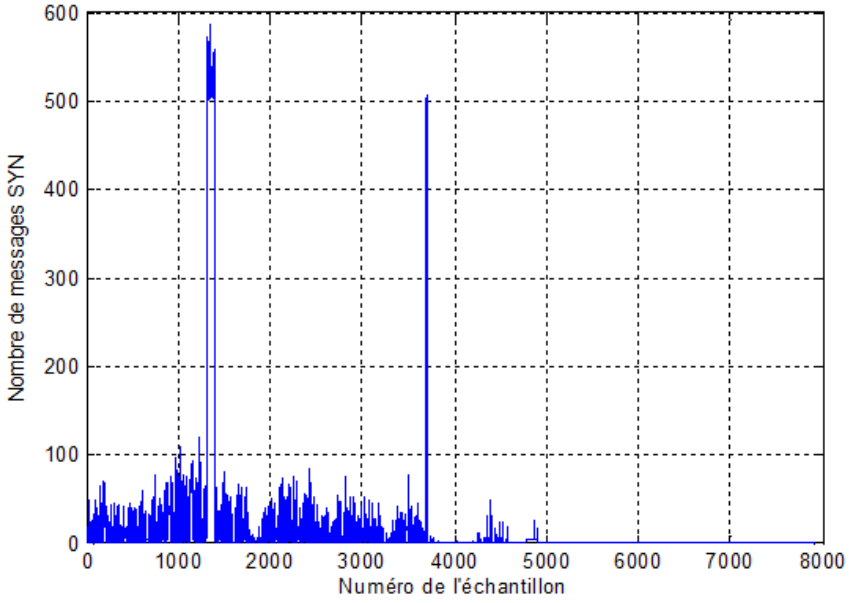

Figure 3.20 : Evolution du nombre de segments SYN en fonction du numéro de l'échantillon (Trafic avec attaques DARPA99 : semaine 5, jour 2)

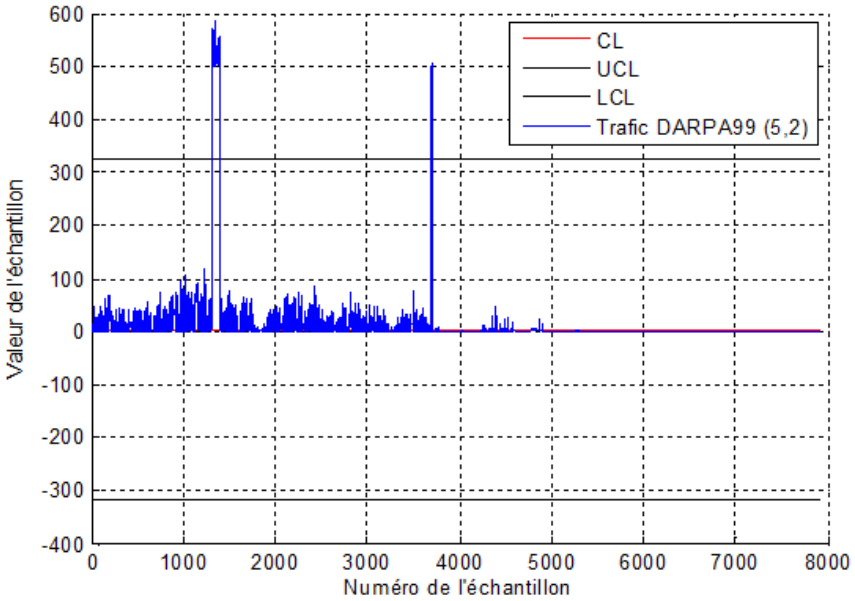

Figure 3.21 : Evolution des valeurs des échantillons en fonction du numéro de l'échantillon (Trafic avec attaques DARPA99 : semaine 5, jour 2)

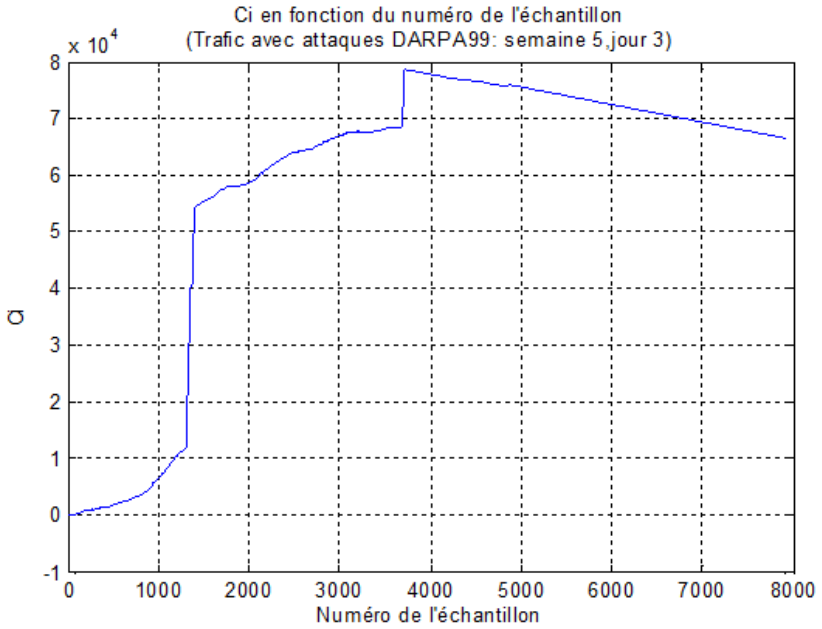

Figure 3.22 : Evolution du Ci en fonction du numéro de l'échantillon

(Trafic avec attaques DARPA99 : semaine 5, jour 2)

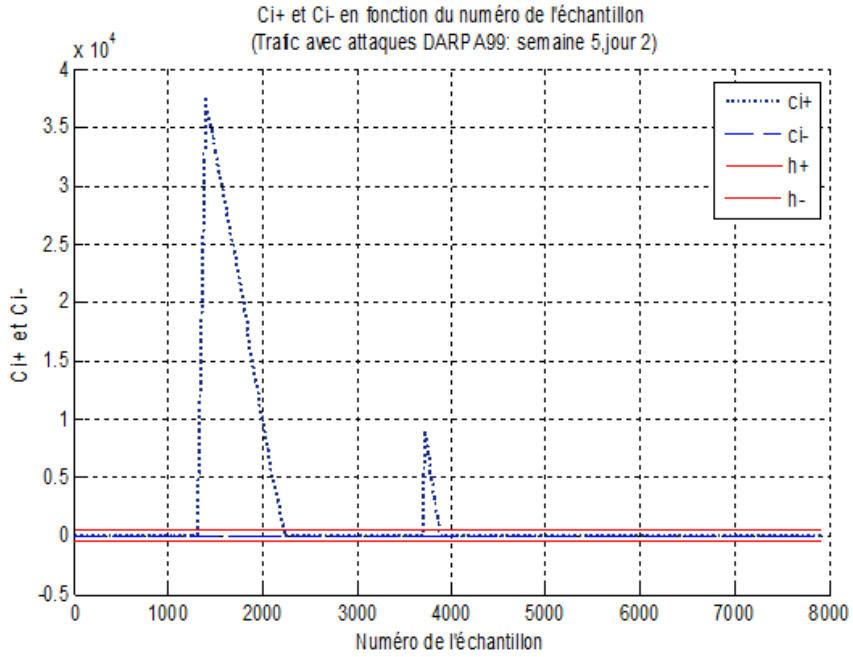

Figure 3.23 : Evolution des C_i^+ et C_i^- en fonction du numéro de l'échantillon

(Trafic avec attaques DARPA99 : semaine 5, jour 2)

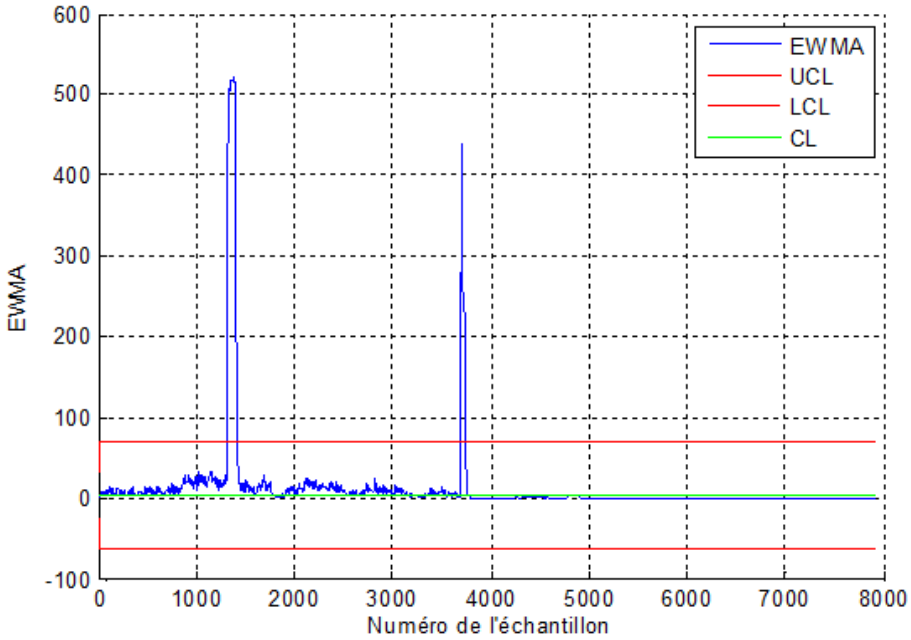

Figure 3.24 : Evolution de l'EWMA en fonction du numéro de l'échantillon (Trafic avec attaques DARPA99 : semaine 5, jour 2)

3.5.2. Détection des attaques SMURF

Dans cette étape, nous utilisons les trois cartes de contrôle pour la détection des attaques SMURF. Nous considérons trois types de trafics anormaux : trafic avec des attaques SMURF de faible intensité, trafic avec attaques SMURF d'intensité élevée et le flux des messages ICMP ECHO REPLY dans le premier jour de la semaine cinq du trafic DARPA99. Le trafic d'apprentissage est celui du jour 3 de la semaine 2. Les limites de contrôle (UCL/LCL) établies avec les trois cartes sont reportées sur le tableau 3.2.

Trafic		Semaine 2, jour 3
Limites de contrôle UCL, LCL	Shewhart	UCL= 20.3568 LCL= -19.0396
	CUSUM	-H= 32.8303 +H= 32.8303
	EWMA	UCL=3.0689/64.047/81.1595/90.3683/95.7832/99.0886/101.1456/102.44/103.26/103.78/104.11 /104,32/104,45 /104,546….. LCL= LCL=3.0689/-57.9/-75.02/-84.23/-89.65/-92.95/-95.007/-96.3/-97.12/-97.64/-97.9711/-98,19/-98, 32/-98,4…..

Tableau 3.2 : Les limites de contrôle UCL et LCL établies avec les cartes Shewhart, CUSUM et EWMA (données d'apprentissage : Semaine 2, jour 3)

Pour étudier le comportement des cartes de contrôle vis-à-vis les attaques de faibles intensité, nous avons généré des attaques SMURF, qui durent 2mn, chaque 3h, pour chaque attaque. La figure 3.25 représente l'évolution des messages ECHO REPLY durant le trafic correspondant. Les figures 3.26, 3.27, 3.28 et 3.29 montrent les résultats de détection obtenus avec les cartes Shewhart, CUSUM et EWMA, respectivement.

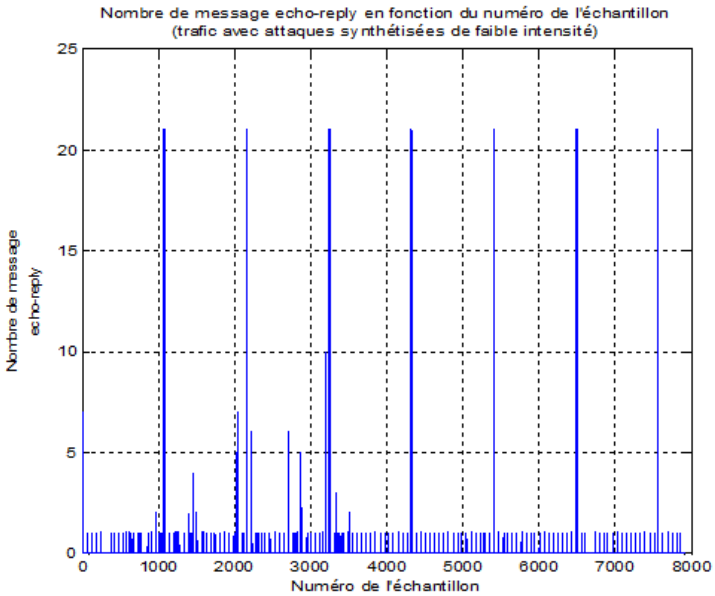

Figure 3.25 : Evolution des messages ECHO REPLYen fonction du nombre de l'échantillon (Trafic avec attaques SMURF de faible intensité)

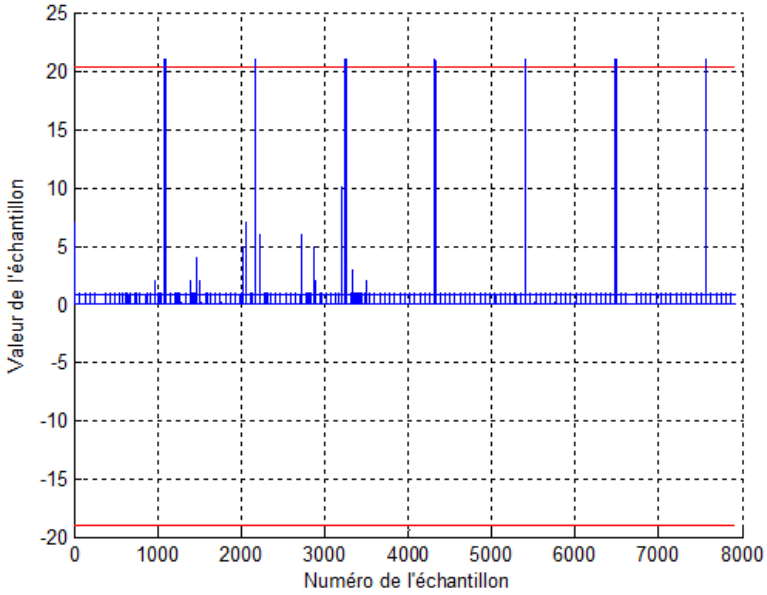

Figure 3.26 : Evolution des valeurs des échantillons en fonction du numéro de l'échantillon (Trafic avec attaques SMURF de faible intensité)

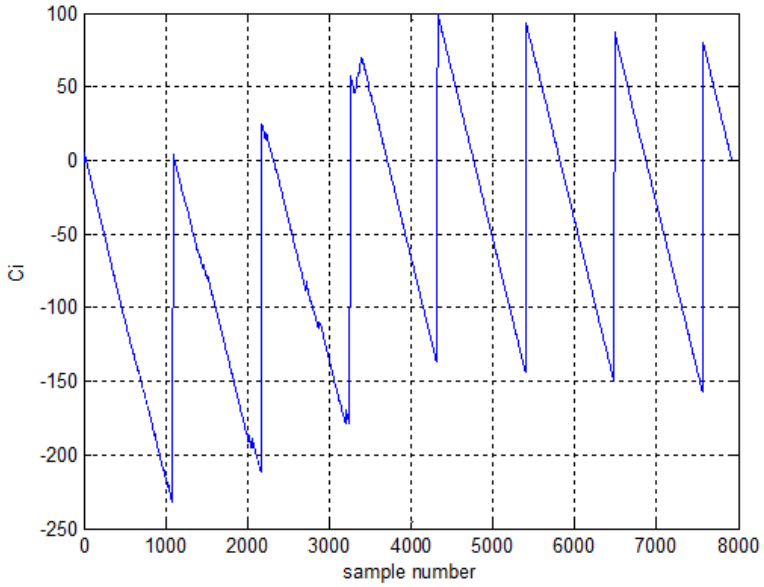

Figure 3.27 : Evolution du Ci en fonction du numéro de l'échantillon (Trafic avec attaques SMURF de faible intensité)

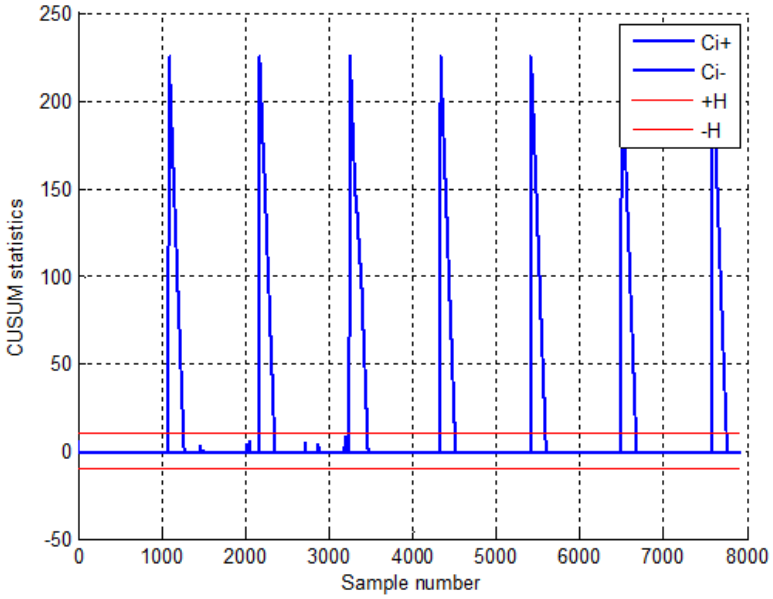

Figure 3.28 : Evolution des C_i^+ et C_i^- en fonction du numéro de l'échantillon (Trafic avec attaques SMURF de faible intensité)

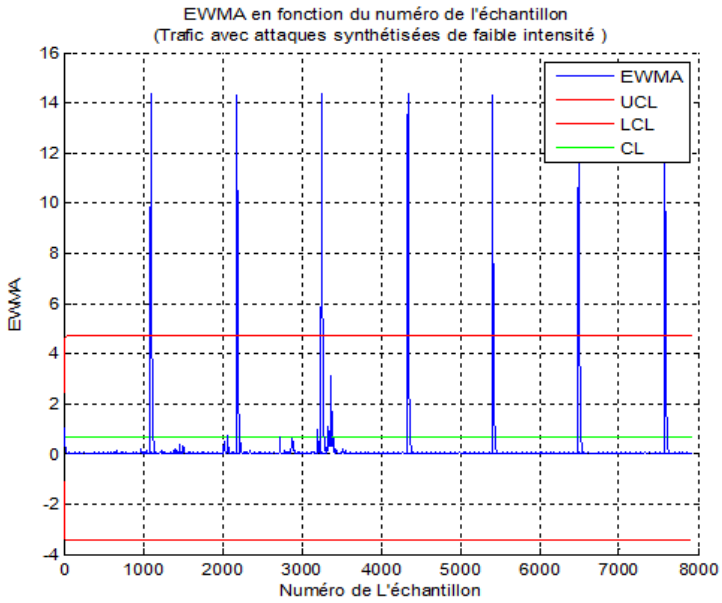

Figure 3.29 : Evolution de l'EWMA en fonction du numéro de l'échantillon (Trafic avec attaques SMURF de faible intensité)

Maintenant, nous étudions le comportement des cartes en vis-à-vis un trafic qui contient des attaques SMURF d'intensité élevée (attaques de 2mn chaque 3 heures). La figure 3. 30 illustre le trafic généré, et les figures 3. 31, 3. 32, 3.33 et 3.34 récapitulent les résultats de détection obtenus avec les trois cartes.

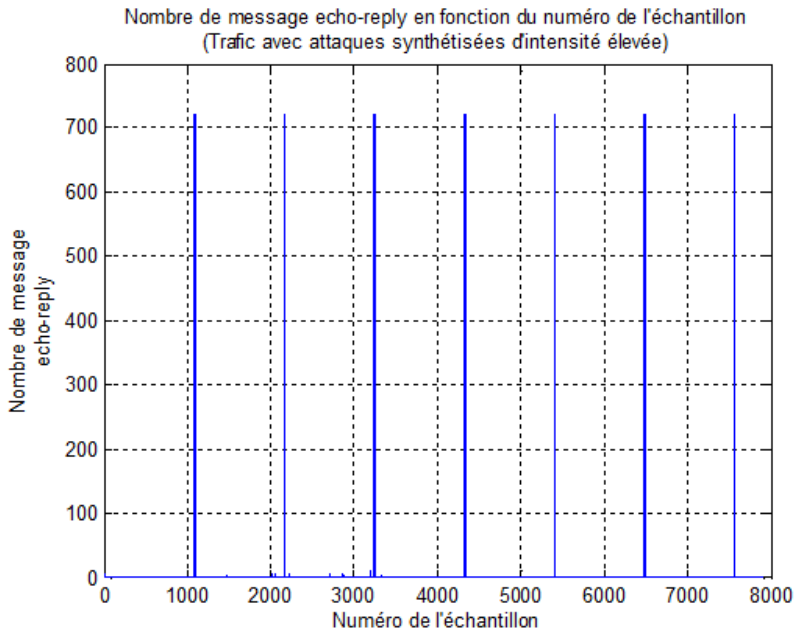

Figure 3.30 : Evolution du nombre de message ECHO-REPLY en fonction du numéro de l'échantillon (trafic avec attaques SMURF d'intensité élevée)

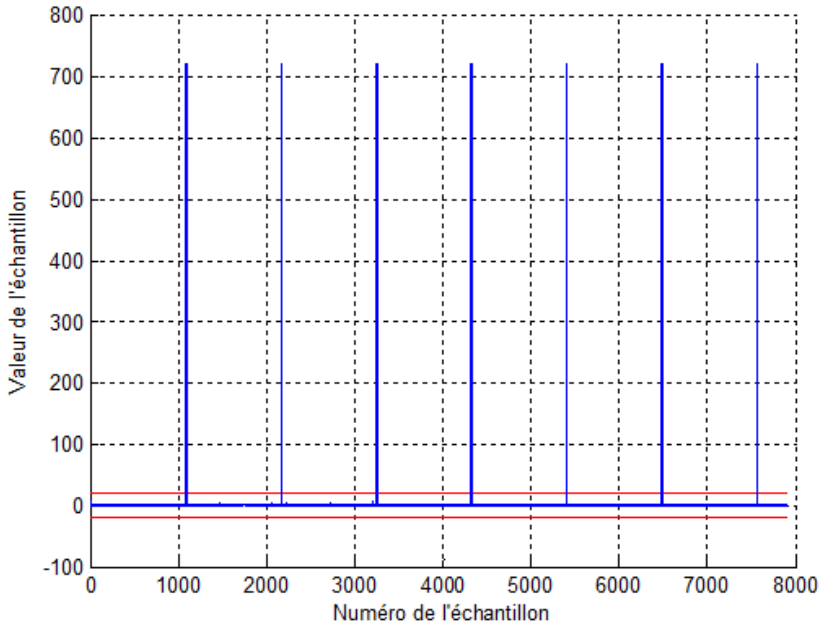

Figure 3.31 : Evolution des valeurs des échantillons en fonction du numéro de l'échantillon (Trafic avec attaques SMURF intensité élevée)

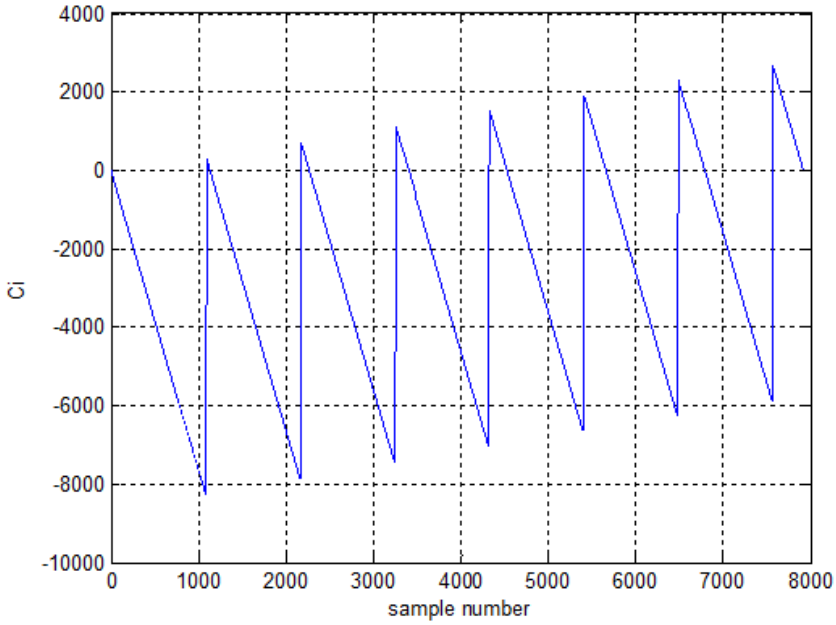

Figure 3.32 : Evolution du Ci en fonction du numéro de l'échantillon

(Trafic avec attaques SMURF d'intensité élevée)

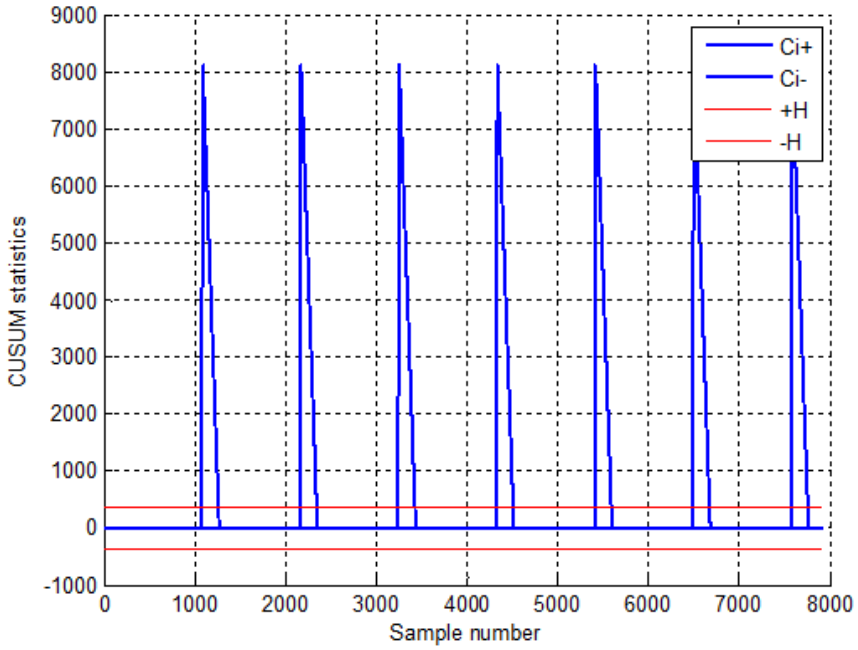

Figure 3.33 : Evolution des C_i^+ et C_i^- en fonction du numéro de l'échantillon

(Trafic avec attaques SMURF d'intensité élevée)

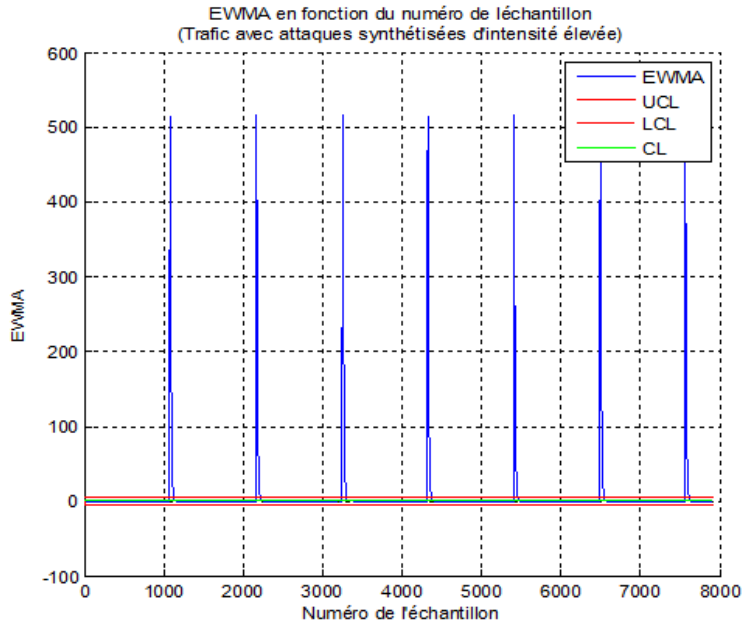

Figure 3.34 : Evolution de l'EWMA en fonction du numéro de l'échantillon (Trafic avec attaques SMURF d'intensité élevée)

Finalement, on applique les trois cartes aux données de test DARPA99 qui contiennent deux attaques SMURF dans la semaine 5 jour 1, une de 2mn, l'autre de 1s. Les figures suivantes présentent ce trafic de test ainsi que les résultats de détection correpondants.

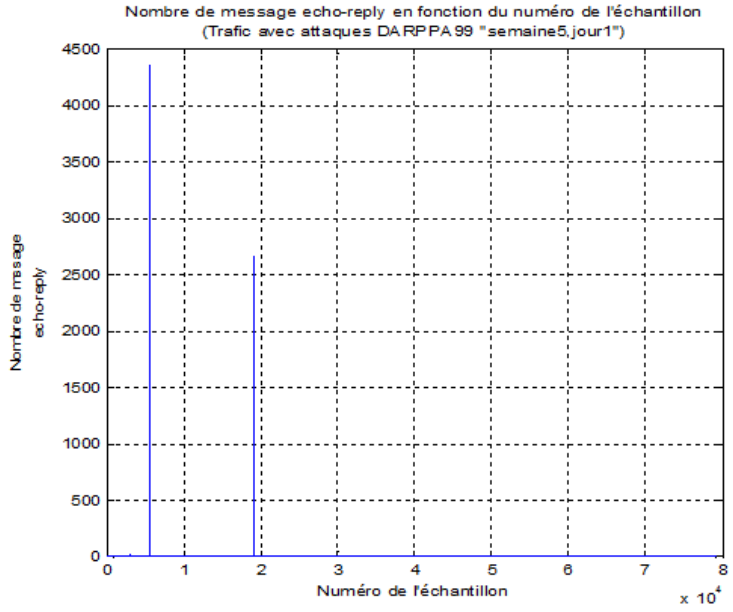

Figure 3.35 : Evolution du nombre de message ECHO-REPLY en fonction du numéro de l'échantillon (Trafic avec attaques SMURF DARPA99 : semaine 5, jour 1)

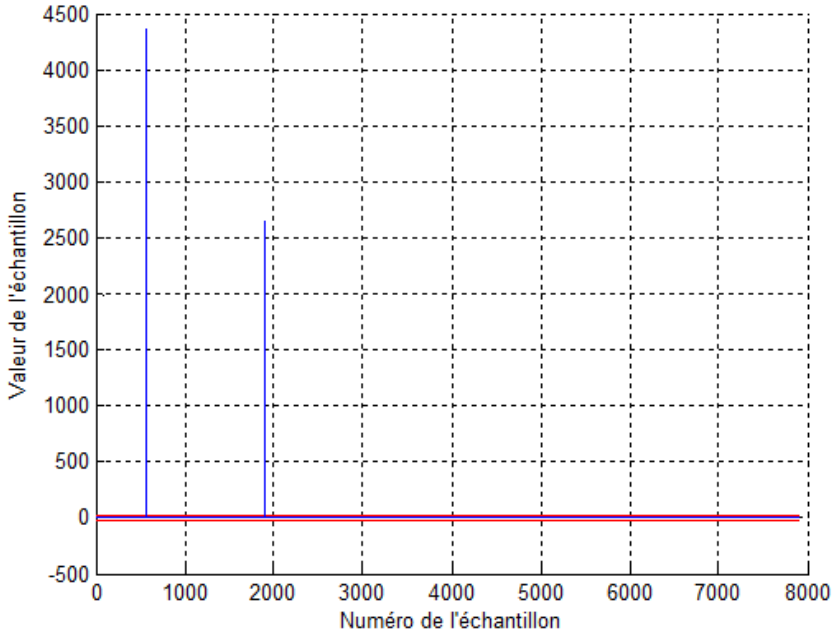

Figure 3.36 : : Evolution des valeurs des échantillons en fonction du numéro de l'échantillon (Trafic avec attaques SMURF DARPA99 : semaine 5, jour 1)

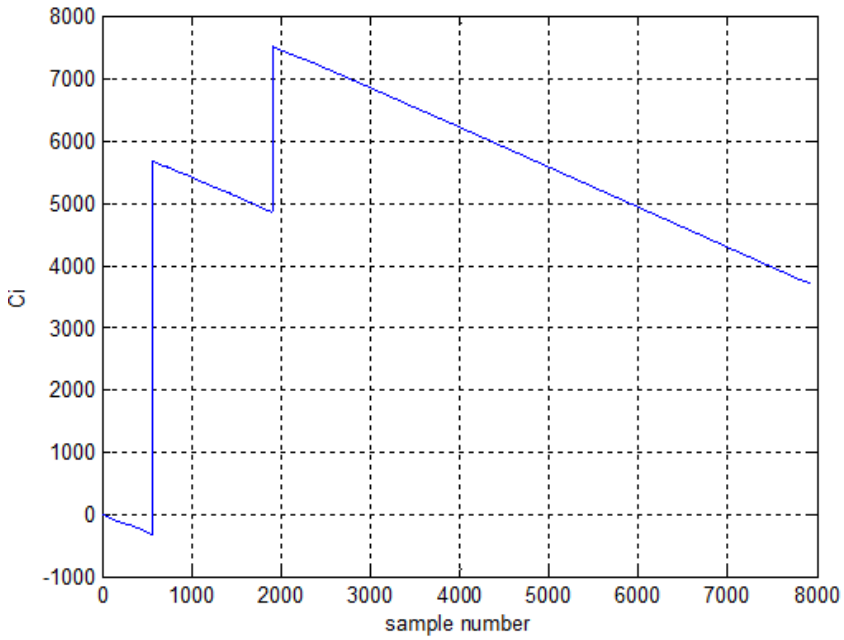

Figure 3.37 : Evolution du Ci en fonction du numéro de l'échantillon (Trafic avec attaques SMURF DARPA99 : semaine 5, jour 1)

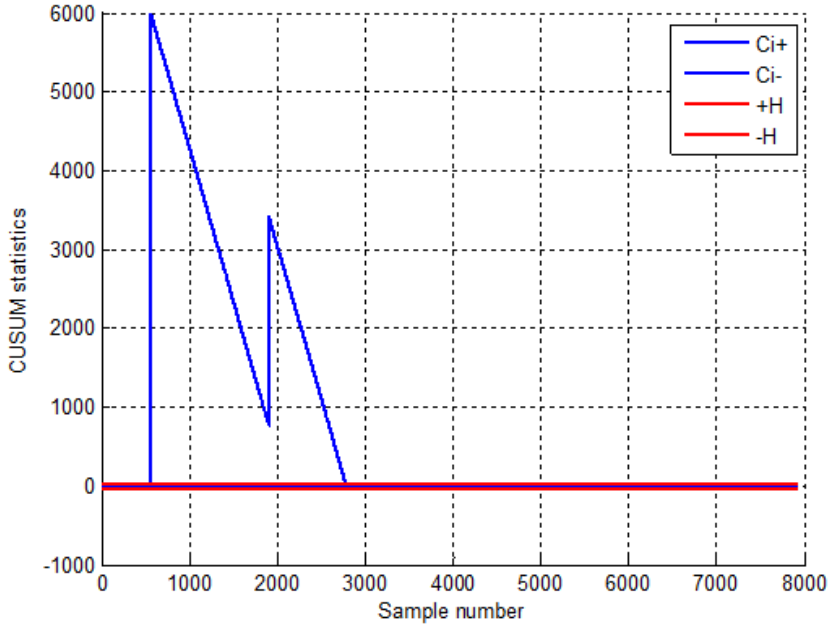

Figure 3.38 : Evolution des C_i^+ et C_i^- en fonction du numéro de l'échantillon

(Trafic avec attaques SMURF DARPA99 : semaine 5, jour 1)

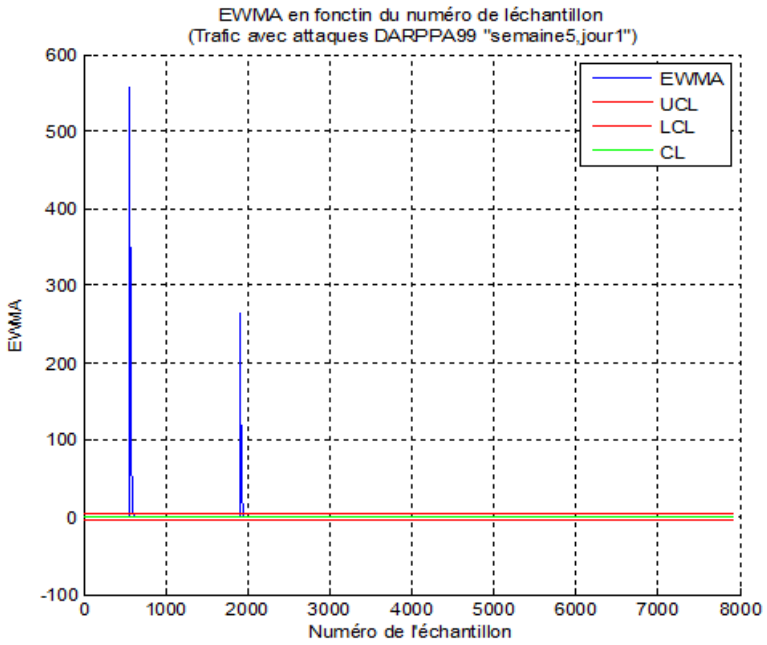

Figure 3.39 : Evolution de l'EWMA en fonction du numéro de l'échantillon

(Trafic avec attaques SMURF DARPA99 : semaine 5, jour1)

3.6. Discussions

Les résultats obtenus montrent que les cartes de contrôle peuvent avoir un intérêt certain dans la détection des attaques DOS/DDOS contre les réseaux IP :

- On remarque que les attaques se manifestent, généralement, par un changement brusque et important des caractéristiques des cartes de contrôle. Lorsque les trafics sont normaux, les paramètres contrôlés dans Shewhart, leurs C_i, C_i^+ et C_i^- dans CUSUM et les statistiques EWMA ont de petites valeurs, et elles deviennent plus importantes durant la présence des différents types d'attaques, reportant ainsi, de grandes déviations du trafic avec attaques DOS/DDOS par rapport à l'état normal du trafic du réseau administré.
- Les trois cartes ont permis la détection totale des attaques avec intensité élevée. Durant les attaques SYN flood et SMURF avec intensité élevée, le nombre des segments TCP SYN et des messages ICMP ECHO REPLY, respectivement, est largement supérieur par rapport l'état normal. Ces écarts importants peuvent être systématiquement révèles par les trois cartes.
- La carte Shewhart n'a pas détecté les attaques de faible intensité. Sachant que ces attaques sont caractérisées par un nombre de paramètres légèrement supérieur par rapport à l'état normal du trafic. Il est difficile de détecter ce type d'attaque en analysant directement ces données. Par contre, avec CUSUM et EWMA, les caractéristiques du trafic contenant ces attaques apparaissent totalement anormales. Considérer l'historique du trafic et l'accumulation des petite déviations a donné des C_i^+ et C_i^- et EWMA nettement supérieures aux limites de contrôles.
- Etant donné que CUSUM est basée sur la somme cumulative des données, on remarque qu'après une attaque, C_i^+et C_i^- prennent plus de temps pour retourner à l'état normal. Ce comportement est totalement inapproprié. En fait la carte annonce de durée supérieure que la durée réelle des attaques, provoquant par conséquence trop de fausses alarmes.
- L'EWMA a permis la détection des attaques de faible intensité. Ces attaques sont caractérisées par un nombre de messages ECHO REPLY légèrement supérieur au nombre de ces messages dans un trafic ICMP normal. Il est difficile de détecter ce

type d'attaque en analysant directement ces données. Par contre, l'EWMA de ces attaques apparait totalement anormal.

- En général, de tous ces résultats, on peut constater que les meilleures performances sont ceux de la carte EWMA. Cette dernière a permis la détection des différents types d'attaques avec leurs différentes intensités, en générant les taux de fausses alarmes les plus faibles par rapport aux cartes Shewhart et CUSUM.
- La détection des attaques avec les trois cartes est basée sur l'hypothèse que les trafics contrôlés possèdent une distribution de probabilité de loi normale. Malheureusement, cette condition n'est pas toujours satisfaite dans la pratique, où généralement les trafics réseau ne peuvent pas être modélisés par une simple distribution Gaussienne. Par conséquence, l'invalidité de l'hypothèse de normalité mène à des résultats de détection trompeuses et qui ne traduisent pas la vraie situation des trafics et des réseaux.

3.7. Conclusion

Les cartes de contrôle sont actuellement largement utilisées dans plusieurs disciplines. Dans ce chapitre nous avons examiné l'utilité de ces cartes dans la détection des anomalies dans un trafic IP, et en particulier celles résultantes des cyber-attaques de types DOS et DDOS. L'idée de base est que les attaques DOS et DDOS provoquent en général une fluctuation d'un ou quelques paramètres du trafic qui circule dans le réseau. Alors elles peuvent être utilisées pour révéler leurs déviations par rapport à l'état normal et reporter les valeurs anormales qui vont au delà des limites de variations normales.

Nous avons utilisé les cartes, Shewhart, CUSUM et EWMA pour détecter les attaques TCP SYN flood et SMURF incluses dans la base de trafic IP DARPA99. Les résultats de détection montrent que ces cartes peuvent avoir un intérêt certain dans ce sens. Il était clair que les attaques étudiées engendrent une augmentation significative des caractéristiques des cartes de contrôle, suffisante pour les détecter. Elles présentent cependant quelques limitations. La carte Shewhart est mal adaptée avec les attaques de faible intensité. Les cartes CUSUM et EWMA peuvent générer de fausses alarmes, en particulier trop avec CUSUM qui prend plus de temps avant de retourner à l'état normal.

Pour exploiter les avantages des cartes de contrôle, et remédier aux limitations, nous proposons une nouvelle vision pour construire ces cartes. Pour réduire la variabilité des paramètres du trafic, nous proposerons les mesures de similarité pour générer les données d'entrée aux cartes de contrôle. Ainsi, pour améliorer les performances des cartes conventionnelles Shewhart et EWMA, nous les combinerons avec ces mesures. Dans le chapitre suivant, nous proposerons de nouvelles techniques de monitoring et de détection d'anomalies de trafic IP à base de la distance CRPS. Les cartes et les mécanismes ainsi proposées seront validées à travers la détection d'une série d'attaques DOS et DDOS dans les environnements IPv4 et IPv6.

Chapitre4

Techniques de détection d'anomalies à base de la distance CRPS

Dans ce chapitre

- La distance CRPS
- Les bases de données de trafics IPv4 et IPv6
- Détection d'anomalies et des Cyber-attaques à base de CRPS
- Résultats et discussions

Chapitre 4

Techniques de détection d'anomalies à base de la distance CRPS

Au fil des ans, plusieurs mécanismes de détection ont été développés pour protéger les réseaux contre différents types d'attaques (les attaques internes et externes).

Le chapitre précédent nous a permis de mettre en évidence l'utilité des cartes de contrôle dans la détection des anomalies de trafic dans les réseaux IP. Nous les avons utilisé pour détecter les attaques DOS et DDOS de types TCP SYN flood et SMURF. Néanmoins, ces cartes, comme de nombreuses techniques de détection à base de détection statistique d'anomalies présentent certaines limitations qui limitent leur déploiement en pratique. Les plus importantes sont : basées sur l'hypothèse de normalité du trafic réseau, utilisent des limites de détection manuelles et mal adaptation à la détection temps réel. Pour résoudre les problèmes liés à ces limitations, nous proposerons dans ce chapitre une nouvelle approche de détection basée sur la distance CRPS. En exploitant les caractéristiques et les avantages du CRPS, notre objectif est de mettre en place des techniques de détection plus adaptées à la détection temps réel, plus flexibles, robustes, ne requière pas une connaissance préalable de la distribution du trafic et qui établaient des limites de contrôle automatiques, adaptables et non-paramétriques. Les mesures CRPS sont appliquées à Shewhart, EWMA et ES pour la détection automatique des anomalies.

Le reste du chapitre est organisé comme suit : on commencera par présenter la distance CRPS. Ensuite, nous introduirons nos motivations et objectifs de l'utilisation de CRPS comme indicateur de détection d'anomalies, en détaillant le principe détection adopté et les mécanismes ainsi développés, à savoir CRPS-Shewhart, CRPS-EWMA et CRPS-ES. On validera, après, les performances de ces mécanismes en présence de plusieurs types d'attaques DOS et DDOS via trois bases de trafic IPv4 et IPv6 (DARPA99, MAWI et trafic ICMPv6). Une comparaison avec des travaux antérieurs est aussi menée.

4.1. La distance CRPS

La distance CRPS a été largement utilisée dans les prévisions probabilistes pour vérifier leur précision [99]. Elle tient en compte à la fois de la netteté et de l'exactitude de ces prévisions. CRPS mesure la différence entre les distributions cumulatives prévues et observées. Elle compare une distribution avec un seul point. Pour une observation x et le CDF (Cumulative Density Function), F d'une variable de prévision probabiliste, la distance CRPS est définie comme [99] [100]:

$$\mathrm{CRPS}(F,x)= \int_{-\infty}^{\infty}\left(F(y)-1\{y\geq x\}\right)^2 dy \tag{4.1}$$

Avec $1\{y \geq x\}$ est la fonction indicative:

$$1(x) = \begin{cases} 0, & x < y \\ 1, & x \geq y \end{cases} \tag{4.2}$$

Une illustration de l'idée de base de la métrique CRPS est donnée par la figure 4.1. La figure 4.1 présente les CDF pour la nouvelle observation et les mesures de prévision. Le CRPS représente la zone de couleur verte.

Lorsque la distribution prédictive de prévision est gaussienne avec moyenne μ et de variance σ^2, CRPS a la formule analytique suivante [100]

$$CRPS(\mathcal{N}(\mu,\sigma^2),x) = \sigma\left[\frac{x-\sigma}{\sigma}\left(2\Phi\left(\frac{x-\sigma}{\sigma}\right)-1\right)+2\phi\left(\frac{x-\sigma}{\sigma}\right)-\frac{1}{\sqrt{\pi}}\right] \tag{4.3}$$

Avec:

ϕ et Φ sont respectivement la fonction de densité de probabilité PDF (Probability Density Function) et la fonction de densité cumulative CDF.

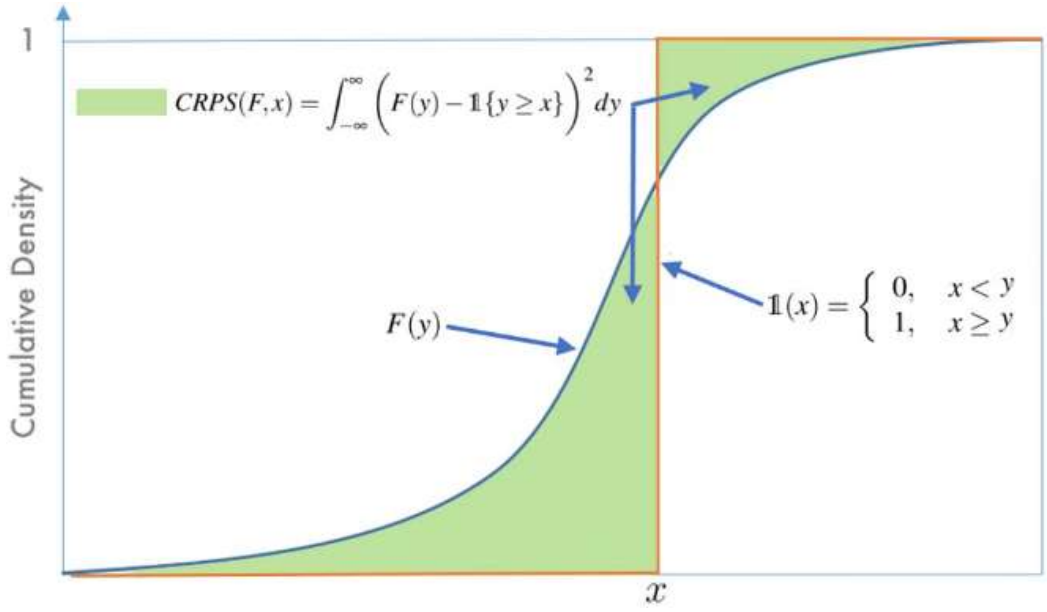

Figure 4.1: Un exemple représentatif de CRPS entre une observation et CDF de données de référence. CRPS (F; x) est la zone délimitée par 1 (x) et F (y).

4.2. Détection d'anomalie à base de CRPS pour la détection des attaques DOS et DDOS

La détection des attaques DOS et DDOS est l'un des défis de sécurité les plus importants auxquels sont confrontées les technologies réseaux. En particulier, les approches de détection des attaques DOS et DDOS basées sur les anomalies ont été partiellement ou totalement conçues à l'aide de mesures basées sur la distribution, telles que la divergence de Kullback-Leibler [101], l'entropie générale [102], HD [103], Chisquare [104] et Rény [30]. Néanmoins, ces méthodes de détection sont généralement basées sur l'hypothèse de normalité de la distribution du trafic réseau. En effet, elles ne fournissent pas de détection appropriée que si les caractéristiques du trafic réseau suivent une distribution normale. Cependant, ces caractéristiques ont généralement une distribution non Gaussienne. De manière générale, la violation de l'hypothèse de normalité peut conduire à un taux élevé de fausses alarmes. Une autre limitation importante concerne le seuil de détection non utilisé ou manuellement établit. En outre, ces techniques sont basées sur des distributions et nécessitent un grand ensemble de données pour détecter correctement les attaques, ce qui les rend inefficaces pour la détection en ligne et rapide, propriété crucial que doivent avoir les solutions de détection des attaques DOS et DDOS.

4.2.1. Motivation et objectifs

Motivés par les performances appropriées de la métrique CRPS et pour résoudre les problèmes susmentionnés, nous avons introduit un mécanisme de détection innovant basé sur la métrique statistique CRPS et le schéma de lissage exponentiel (ES) [105] pour permettre une détection efficace des attaques DOS et DDOS .

Nous avons conçu le mécanisme CRPS-ES avec les caractéristiques suivantes:

- Il n'y a aucune hypothèse concernant la normalité des données et une estimation plus réaliste de la distribution du trafic est proposée. Plus précisément, nous avons appliqué l'estimation de la densité du noyau (KDE) pour estimer de manière non-paramétrique la distribution sous-jacente à la statistique CRPS-ES.
- La détection des attaques est basée sur un seuil automatique et non-paramétrique. Ici, ES est appliqué aux mesures CRPS pour définir le seuil de détection et distinguer les attaques. Le schéma CRPS-ES prend en compte les données passées dans les statistiques de détection, ce qui le rend sensible aux attaques naissantes. Des attaques sont déclarées si les mesures CRPS-ES entre le trafic de test et le trafic normal dépassent le seuil de détection.
- CRPS-ES est plus approprié pour la détection en temps réel. Contrairement aux techniques susmentionnées, dans le mécanisme CRPS-ES, seule la nouvelle mesure du trafic est comparée à la distribution du trafic sans attaque, ce qui la rend appropriée pour la surveillance en temps réel.
- De plus, CRPS-ES est indépendant du protocole et du réseau. Cela signifie qu'il peut être appliqué pour détecter différents types d'attaques DOS et DDOS et dans toutes sortes de réseaux. Différents flux peuvent être séparés et surveillés indépendamment selon les messages et protocoles qu'ils incluent.

En outre, pour améliorer les performances des cartes de contrôle conventionnelles Shewhart et EWMA, nous les avons intégrer avec la distance CRPS. Dans les deux nouvelles cartes CRPS-Shewhart et CRPS-EWMA, les statistiques de décision et les limites de contrôles dérivées sont calculées à partir des mesures CRPS et non pas directement aux paramètres de trafic contrôlé.

4.2.2. Principe de détection d'anomalies par CRPS pour la détection des attaques DOD/DDOS

La distance CRPS est très sensible aux changements, c à d, elle permet de révéler même les petites déviations des mesures contrôlées par rapport aux mesures de référence. En effet, les valeurs de CRPS tendent vers 0 lors ces mesures coïncident avec les mesures de référence. Dans le cas contraire, les valeurs de CRPS augmentent proportionnellement avec le degré de divergence entre les deux ensembles de mesures. Nous exploitons ce comportement pour la détection des anomalies de trafic.

La caractéristique souhaitable du CRPS est sa capacité de comparer une distribution complète avec une (seule) observation qui le rend appropriée pour la détection d'anomalies en ligne (en temps réel). Fondamentalement, pour détecter les attaques DOS et DDOS à l'aide de la métrique CRPS, chaque nouvelle mesure du trafic réseau sera comparée avec la distribution de trafic sans attaques de référence. En raison de sa sensibilité aux changements, le CRPS est approprié pour quantifier la déviation des attaques par rapport au trafic normal. De grandes valeurs de CRPS reflètent la présence d'attaques potentielles dans le trafic surveillé.

En fait, CRPS avec de petites valeurs proches de zéro fait référence à un trafic normal. Cependant, il est clair que la métrique CRPS augmente de manière significative lors des attaques. Cela rend la métrique CRPS utile comme indicateur pour détecter les flux de trafics malveillants.

4.2.3. Cartes de contrôle paramétriques à base de CRPS

Pour développer des techniques de détection d'anomalies efficaces avec une sensibilité plus élevée, nous avons intégré la mesure CRPS avec les deux cartes de contrôle, Shewhart et EWMA.

- **Carte de contrôle CRPS-Shewhart**

Combine la mesure CRPS et la carte contrôle Shewhart. Plus précisément, dans la carte CRPS-Shewhart, les mesures CRPS sont utilisées comme entrée de la carte Shewhart pour la détection des anomalies. Lorsque la $i^{ème}$ valeur CRPS est au-delà des limites de contrôle, alors nous affirmons qu'il y a une anomalie. Sinon, le réseau supervisé est considéré sous contrôle. Les limites de contrôle sont établies comme suit [106]:

$$UCL_s, LCL_s = \mu_0^{CRPS} \pm 3\sigma_0^{CRPS} \qquad (4.4)$$

Où μ_0^{CRPS} et σ_0^{CRPS} sont la moyenne et l'écart type des mesures CRPS dans l'état sans anomalies. UCLs et LCLs représentent les limites de contrôle utilisées pour détecter les anomalies.

- **La carte de contrôle CRPS-EWMA**

Dans la carte CRPS-EWMA, EWMA est appliquée aux mesures CRPS pour la détection des anomalies. La statistique EWMA est calculée comme suit [107]:

$$z_i^{CRPS} = \lambda CRPS_i + (1-\lambda) z_i^{CRPS} \qquad (4.5)$$

Avec:

$CRPS_i$: est la mesure CRPS actuelle

z_0^{CRPS} : représente la moyenne μ_0^{CRPS} des mesures CRPS sans anomalies.

Les limites de contrôle de CRPS-EWMA sont définies comme :

$$UCL_s, LCL_s = \mu_0^{CRPS} \pm L\sigma_0^{CRPS} \sqrt{\left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2i}]} \qquad (4.6)$$

- **Détection des attaques DOS et DDOS avec CRPS-Shewhart et CRPS-EWMA**

(1) Nous calculons les limites de contrôle de chaque carte en utilisant les données d'apprentissage.

(2) Pour les paramètres d'adjustement de EWMA, nous avons utilisé les choix populaires: L = 2. 7 et λ = 0.1

(3) Nous calculons les statistiques CRPS-Shewhart et CRPS-EWMA pour les données de test.

(4) Si les statistiques obtenues dépassent les limites de contrôle, alors une attaque est déclarée.

4.2.4. Carte de contrôle non-paramétrique à base de CRPS : CRPS-ES

Pour améliorer l'efficacité de détection, une procédure de lissage exponentiel (ES) est appliquée aux mesures CRPS. La principale raison du lissage exponentiel des mesures CRPS (CRPS-ES) est d'inclure des informations des mesures précédentes et actuelles dans le processus de décision, ce qui le rend efficace pour découvrir de petites anomalies.

Le schéma de lissage exponentiel (ES) est appliqué aux mesures CRPS pour établir un seuil de décision et signaler la présence de trafics d'attaques. En d'autres termes, dans le mécanisme CRPS-ES proposé, les séquences CRPS sont lissées exponentiellement pour améliorer encore sa sensibilité aux événements anormaux (attaques). Pour ce faire, en fonction de l'attaque ciblée, les paramètres du trafic (par exemple, le nombre d'octets, le nombre de paquets et l'adresse IP) sont utilisés comme variable d'entrée du mécanisme CRPS-ES. Plus précisément, pour détecter les attaques TCP SYN flood, SMURF et ICMPv6, nous avons appliqué CRPS-ES aux nombre de segments SYN, de messages ICMPv4 ECHO-REPLY et de messages ICMPv6 reçus par temps échantillonnage, respectivement.

Définissons la séquence de mesures CRPS calculée en (4.1): CRPS = [d1… dn]. La statistique CRPS-ES est calculée comme suit:

$$z_t^{CRPS} = \upsilon d_t + (1-\upsilon) z_{t-1}^{CRPS} \qquad (4.7)$$

Où la valeur initiale Z_0^{CRPS} est la moyenne μ_0^{CRPS} du vecteur CRPS sans anomalies.

υ ($0 < \upsilon \leqslant 1$) est une paramètre d'ajustement.

Le seuil de détection d'anomalies CRPS-ES est conçu comme [107],

$$h_{ES} = \mu_0^{CRPS} + L\sigma_0^{CRPS} \sqrt{\frac{\nu}{(2-\upsilon)}[1-(1-\nu)^{2t}]} \qquad (4.8)$$

où L désigne la largeur du seuil de décision.

Il convient de noter que le seuil paramétrique dans le schéma CRPS-ES est calculé sur la base de l'hypothèse de normalité des données de trafic. Cependant, l'hypothèse de normalité des données de trafic n'est pas valable principalement en raison des caractéristiques dynamiques du trafic réseau. Lorsque l'hypothèse Gaussienne n'est pas vérifiée, les résultats de la détection seraient inappropriés. Pour atténuer ce problème, la distribution de la statistique CRPS-ES pourrait être estimée en utilisant l'estimation de densité de noyau KDE [108], qui est une approche d'estimation de densité de probabilité non-paramétrique. Dans cette approche, tout d'abord, la distribution de la statistique CRPS-ES dans l'équation (4.7) est estimée via KDE en utilisant des données sans attaques. Autour d'un point xi, KDE est formulée comme suit [108]:

$$\hat{f}(x) = \frac{1}{nH} K \frac{(x-x_i)}{H} \qquad (4.9)$$

Où K est la fonction du noyau, ici le noyau gaussien est utilisé.

n est le nombre de mesures.

x est le point de données considéré.

H est la largeur de bande de lissage du noyau qui détermine la qualité d'estimation; sa valeur optimale peut être calculée comme suit [109]:

$$H = 1.06\sigma n^{-0.2} \quad (4.10)$$

Ensuite, le seuil non-paramétrique du mécanisme CRPS-ES est défini comme le (1-α) -ième quantile de la distribution estimée. Une anomalie est signalée lorsque la statistique CRPS-ES dépasse ce seuil de décision.

Le schéma CRPS-ES conçu peut être récapitulé comme suivant :

- **Etape 1:** pour chaque observation xi dans l'ensemble de données de test, calculer les séquences CRPS.
- **Etape 2:** calculer les séquences CRPS-ES en utilisant l'équation (4.7).
- **Etape 3:** Estimer la distribution sous-jacente aux séquences CRPS-ES via KDE.
- **Etape 4**: A partir de la distribution de CRPS-ES, déterminer le seuil de détection de manière non-paramétrique comme le (1-α)-ième quantile de la distribution estimée des statistiques CRPS-ES calculées par KDE.
- **Etape 5:** une attaque est déclarée lorsque la statistique CRPS-ES dépasse le seuil de détection.

Dans l'ensemble, dans la stratégie de détection d'anomalies basée sur CRPS-ES, tout d'abord, différentes paramètres (par exemple, le segment TCP, l'adresse IP et les messages ICMP) sont extraites du trafic réseau collecté en fonction des attaques ciblées. Ensuite, les données d'apprentissage sans attaques sont utilisées comme entrée de la métrique CRPS lissée exponentiellement CRPS-ES. Le seuil de détection est calculé de manière non-paramétrique sur la base de la distribution CRPS-ES estimée à l'aide de KDE. Ensuite, pour révéler les attaques DOS et DDOS qui pourraient être présentes dans le trafic testé, les valeurs CRPS-ES sont continuellement comparées au seuil de détection calculé précédemment. Le trafic est normal pour toutes les observations détectées par le CRPS-ES jusqu'à sa limite de détection. Un trafic anormal et éventuellement la présence d'attaques DOS et DDOS sont tous des observations détectées au-dessus du seuil CRPS-ES. Le schéma conceptuel du mécanisme CRPS-ES proposé est illustré sur la figure 4.2.

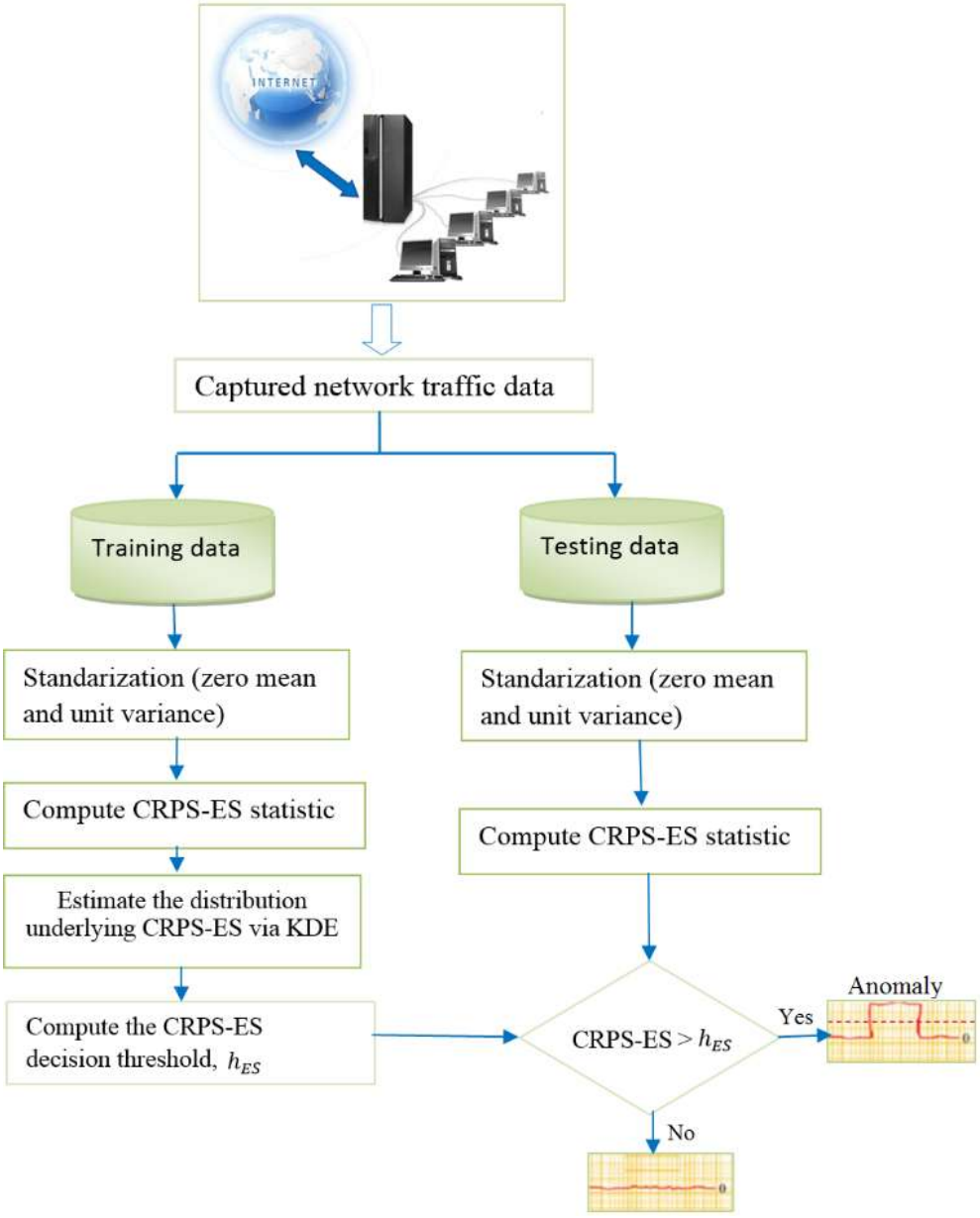

Figure 4.2: Schéma conceptuel du mécanisme CRPS-ES

4.3. Bases de trafics

Les mécanismes de détection basés sur CRPS seront utilisés pour détecter les attaques DOS et DDOS. Plus précisément, nous étudions leur capacité à détecter les types d'attaques DOS et DDOS les plus populaires, à savoir TCP SYN flood, UDP flood, l'attaque SMURF et les attaques par inondation basées sur le protocole ICMPv6. Pour cela, nous utiliserons les traces de trafics collectées à partir des bases de données DARPA99, MAWI et ICMPv6.

- **La base DARPA99 :** la description de cette base ainsi que des attaques associées sont présentées dans le chapitre 3 (cf. paragraphe 3.4).

- **La base MAWI (Measurement and Analysis on the WIDE Internet) :** la base MAWI est un trafic Internet réel fourni par le référentiel de trafic du MAWI Working Group. Dans cette base de données, le trafic est capturé à partir de nombreuses liaisons transpacifiques (c'est-à-dire, sampleponit-A, sampleponit-C, sampleponit-D et sampleponit-F) entre le réseau Japonais WIDE et les Etats-Unis. L'échantillon point-F, qui est le plus utilisé, fournit une trace quotidienne de 15mn. Nous avons utilisé la trace TCPDUMP du 1er janvier 2010; de 14h00 à 14h15mn [110].
- **ICMPv6 traffic dataset** [111]: en utilisant le réseau de la figure 4.3, une base de données de trafic ICMPv6 est générée. Cette base contient deux types de trafic : le trafic normal sans attaques et le trafic anormal qui introduit plusieurs attaques DOS basées sur le protocole ICMPv6. Les deux types de trafic sont capturés par la station Monitor en utilisant le logiciel Wireshark. Le trafic normal ou le trafic d'apprentissage représente 48 h de trafic ICMPv6 généré par les différentes stations pour assurer le bon fonctionnement du réseau. Le trafic anormal ou le trafic de test est un trafic ICMPv6 qui contient de nombreuses attaques DOS (Ping flood, NA flood, NS flood…). Ces attaques sont générées par le THC-toolkit.

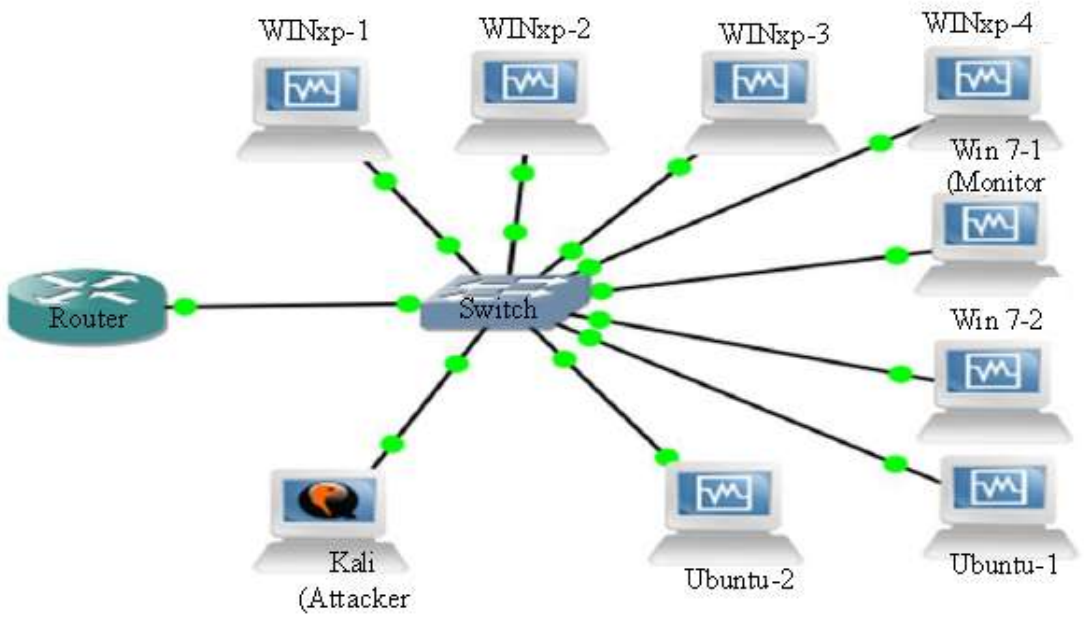

Figure 4.3: Topologie du réseau utilisé pour générer le trafic ICMPv6 [111]

Pour extraire les paramètres de détection utilisés pour la détection des différents types des attaques DOS et DDOS, la même procédure de prétraitement décrite dans le chapitre 3 (cf. paragraphe 3.4) est appliquée sur les trois bases de trafic.

4.4. Résultats de détection

Les mécanismes de détection conçus à base de CRPS seront utilisés pour le monitoring des attaques DOS et DDOS. Notez que ces types d'attaques sont toujours effectués contre la majorité des réseaux actuels et il devrait persister avec les futures technologies.

4.4.1. Performances des cartes de contrôle paramétriques CRPS-Shewhart et CRPS-EWMA

Pour évaluer les performances des cartes paramétriques CRPS-Shewhart et CRPS-EWMA, nous avons construit une nouvelle base de trafic à partir des données DARPA 99, contenant uniquement des segments SYN (la base DARPA99 / SYN). En utilisant ces données, nous avons généré également des attaques par TCP SYN flood intermittentes, de différentes intensités.

❶ Attaques TCP SYN flood intermittentes avec différentes intensités

Dans ce scénario, nous étudions la capacité des approches CRPS-Shewhart et CRPS-EWMA à détecter les attaques TCP SYN flood intermittentes avec différentes intensités. Ces attaques se produisent et disparaissent à plusieurs reprises. Ici, toutes les trois heures, nous introduisons dix minutes d'attaques TCP SYN flood.

Les résultats de détection obtenus avec les quatre cartes Shewhart, EWMA, CRPS-Shewhart et CRPS-EWMA sont présentés sur les figures 4.4, 4.5, 4.6 et 4.7, respectivement. Les cartes peuvent détecter ces attaques avec des intensités modérées. Les figures 4.4 et 4.7 montrent que les Shewhart et CRPS-Shewhart détectent ces attaques mais en générant plusieurs fausses alarmes. D'autre part, la carte CRPS-EWMA détecte ces attaques sans fausses alarmes.

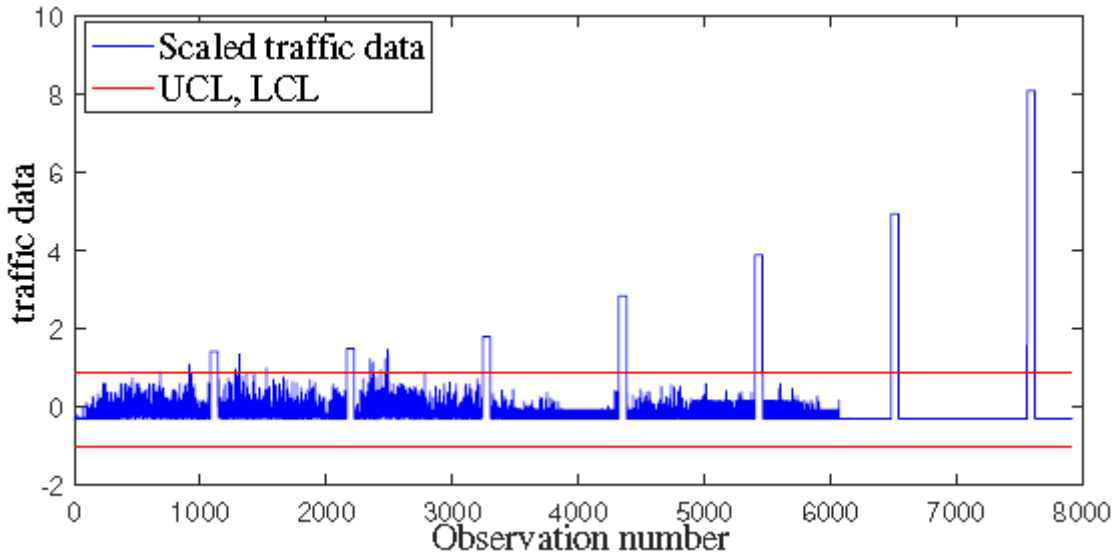

Figure 4.4: Détection des attaques TCP SYN flood intermittentes par la carte Shewhart

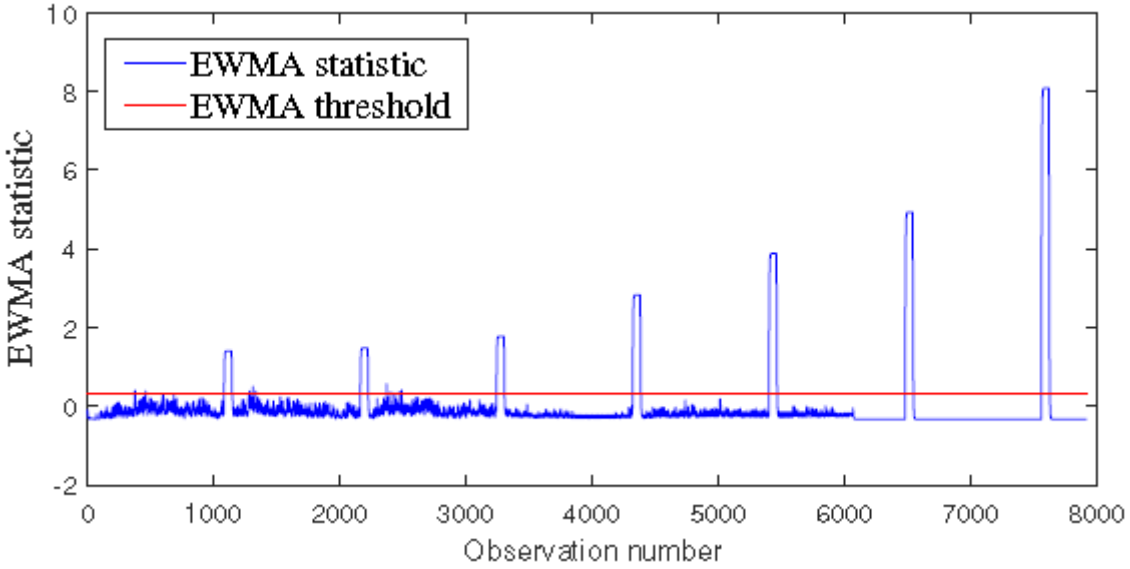

Figure 4.5: Détection des attaques TCP SYN flood intermittentes par la carte EWMA

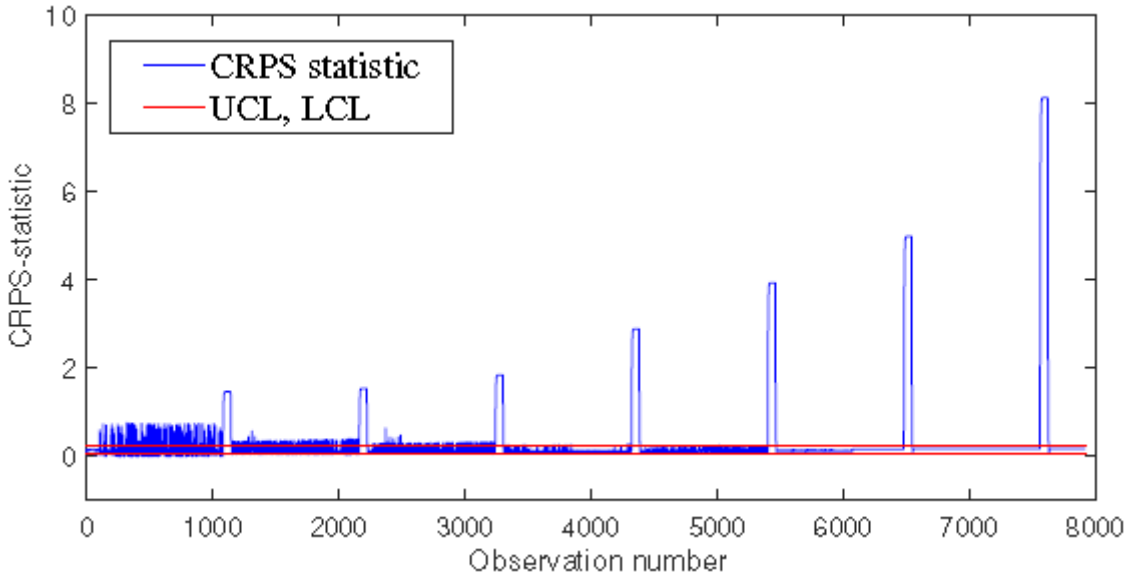

Figure 4.6: Détection des attaques TCP SYN flood intermittentes par la carte CRPS-Shewhart

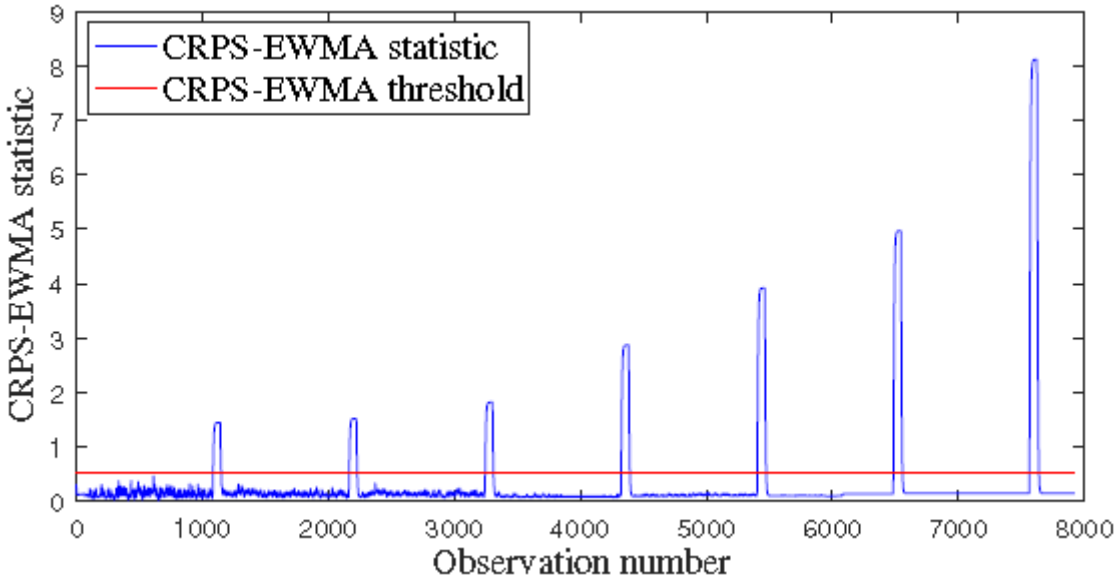

Figure 4.7: Détection des attaques TCP SYN flood intermittentes par la carte CRPS-EWMA

❷ Attaques TCP SYN flood du trafic DARPA (semaine 5, jour 2)

Dans ce scénario, nous évaluons les performances de CRPS-Shewhart et CRPS-EWMA dans la détection des attaques TCP SYN flood inclues dans la semaine 5, jour 2 de la base DARPA 99 [98]. Le trafic correspondant comprend deux attaques. La première commence à 11h38mn04s contre Marx (@: 172.16.114.50) avec une durée de 13mn41s. La deuxième était à 18h16mn05s contre le routeur (@IP: 192.168.1.1) pendant 3mn26s. Les

résultats de détection sont présentés sur les figures 4.8, 4.9, 4.10 et 4.11. Ces résultats montrent que la méthode proposée CRPS-EWMA donne les meilleurs performances que Shewhart, EWMA et CRPS-Shewhart et a permis la plus grande sensibilité. En fait, CRPS-EWMA détecte correctement ces attaques sans fausses alarmes.

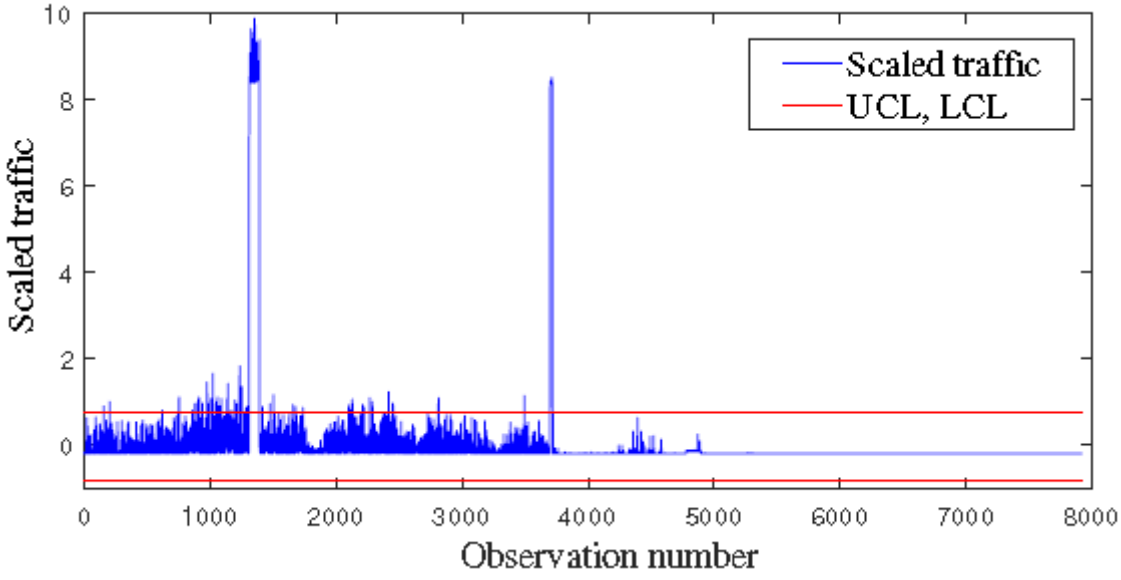

Figure 4.8: Détection des attaques TCP SYN flood dans le trafic DARPA99 (semaine 5, jour 2) par la carte Shewhart

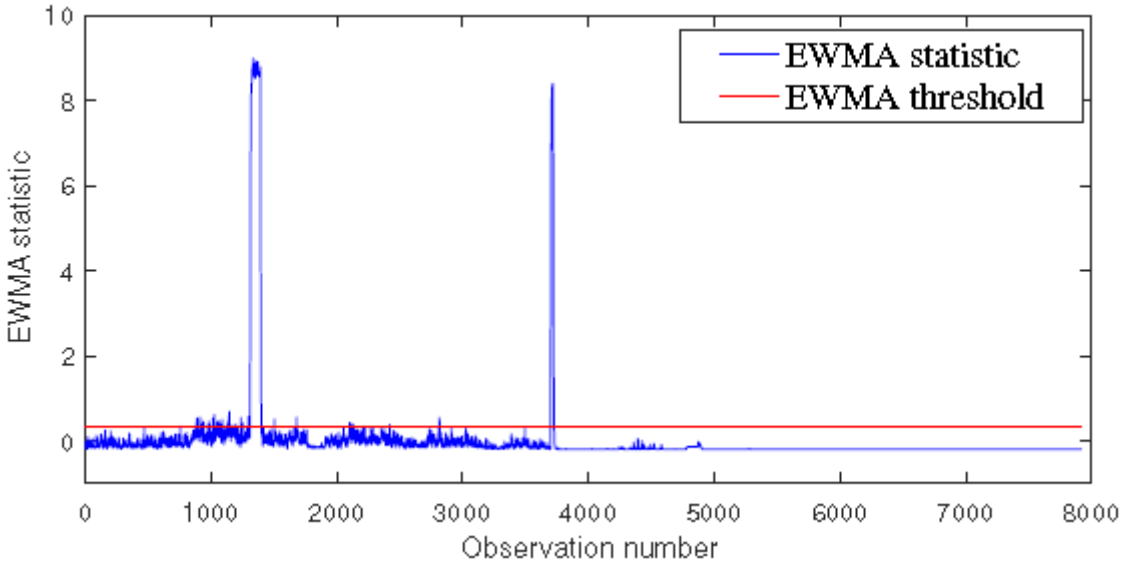

Figure 4.9: Détection des attaques TCP SYN flood dans le trafic DARPA99 (semaine 5, jour 2) par la carte EWMA

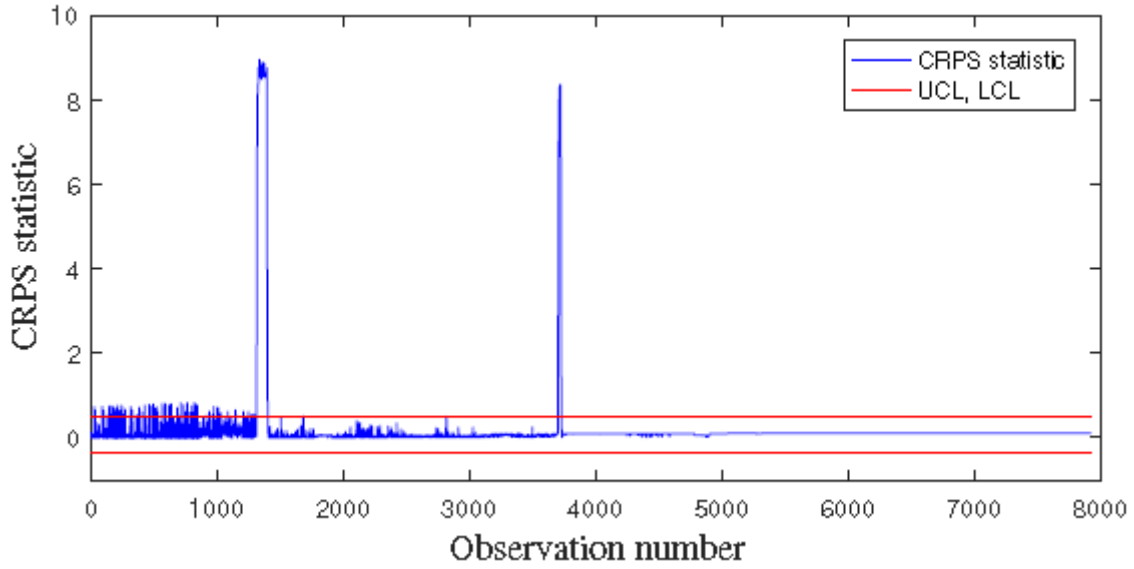

Figure 4.10: Détection des attaques TCP SYN flood dans le trafic DARPA99 (semaine 5, jour 2) par la carte CRPS-Shewhart

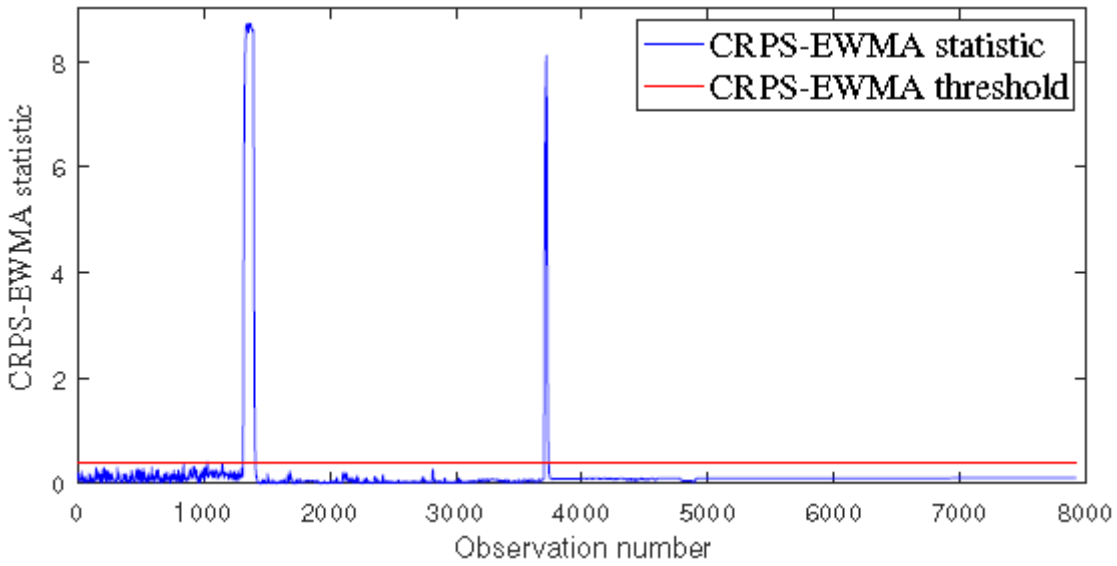

Figure 4.11: Détection des attaques TCP SYN flood dans le trafic DARPA99 (semaine 5, jour 2) par la carte CRPS-EWMA

4.4.2. Performances de la carte de contrôle non-paramétrique CRPS-ES

Nous évaluons, ici, les performances de CRPS-ES en utilisant les trois bases de trafic, DARPA99, MAWI et ICMPv6.

4.4.2.1. Résultats de détection avec la base DARPA99

Dans cette partie, la capacité de détection du CRPS-ES est vérifiée à travers la détection des attaques TCP SYN flood et les attaques SMURF dans le trafic réseau de la base DARPA99.

Dans le scénario d'une attaque par TCP SYN flood, nous considérons trois attaques. La première attaque a eu lieu dans la semaine 5, jour 1 à 18h04mn04s et dure 6mn51s. La seconde survient dans la semaine 5, jour 2, commence à 11h38mn04s et avec une durée de 13mn41s. La troisième a apparaissait également dans la semaine 5, jour 2 à 18h16mn05s pendant 3mn26s.

Le deuxième scénario concerne les attaques SMURF. Plus précisément, nous étudions cinq attaques ICMP SMURF, qui ont ciblé la même victime au cours des quatrième et cinquième semaines. Le début était avec la semaine 4, au cours de laquelle deux attaques de 1s ont été lancées le jour 1 à 21:34:16 pm et 21:34:26 pm, une attaque de 1s à 18:29:25 pm au jour 3 et une autre attaque de 2s à 08:45:18 le jour 5. Enfin, la victime a été de nouveau attaquée au cours de la semaine 5, jour 1 à 09:33:00 pendant 2mn. Nous baptisons toutes ces attaques par « attaque WiDi » (c à d attaque dans la semaine i et le jour i).

Les caractéristiques générales de ces attaques sont récapitulées dans le tableau 4.1. En revanche, les données normales ou sans attaques représentent environ 1320 enregistrements correspondant à 22 h du trafic réseau.

Attaque		Semaine	Jour	Temps d'apparition	Durée
TCP SYN flood	Attaque 1	5	1	18:04:04	6mn51s
	Attaque 2	5	2	11:38:04	13mn41s
	Attaque 3	5	2	18:16:05	3mn26s
SMURF	Attaque 1	4	1	21:34:16	1s
	Attaque 2	4	1	21:34:26	1s
	Attaque 3	4	3	18:29:25	1s
	Attaque 4	4	5	08:45:18	2s
	Attaque 5	5	1	09:33:00	2mn

Tableau 4.1: Caractéristiques des attaques TCP SYN flood et SMURF dans la base DARPA99

Pour détecter les attaques TCP SYN flood, le flux de segments SYN reçus par la victime est contrôlé. La figure 4.12 illustre la détection basée sur le trafic W5D1. Les résultats montrent que l'attaque s'est produite à partir des enregistrements 605 à 611 avec un taux de 2928 segments SYN par temps d'observation. Il est clairement illustré que lorsque le trafic surveillé (c'est-à-dire le trafic W5D1) est exempt d'attaques, les mesures CRPS tombent sous le seuil de détection (hes = 0,7). Cela signifie que tel trafic a un comportement similaire au trafic normal sans attaques. D'autre part, de grandes valeurs de statistiques CRPS sont obtenues lorsque le trafic inspecté comprend des attaques TCP SYN flood qui sont à l'origine des attaques détectées. Au cours de ces attaques, la statistique CRPS était d'environ 12, dépassant largement le seuil hes.

La figure 4.13 montre la capacité de détection du mécanisme CRPS-ES en présence des deux attaques survenues dans le trafic W5D2. Lors de la première attaque, la victime a été submergée par une moyenne de 3027 SYN/temps d'observation, et lors de la deuxième attaque, la victime a été inondée avec 10256 segments SYN à chaque instant. Dans ce trafic, le seuil de détection hes = 0,51, or les deux attaques TCP SYN flood ont augmenté les statistiques CRPS à 9,1 et 5,3, respectivement.

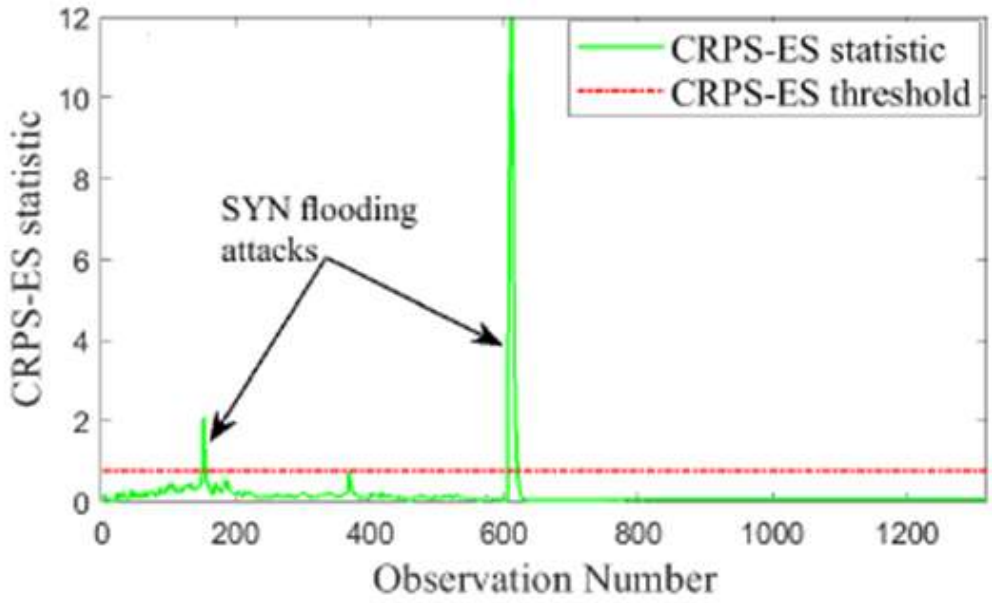

Figure 4.12 : Résultat de détection des attaques TCP SYN flood dans W5D1 par CRPS-ES

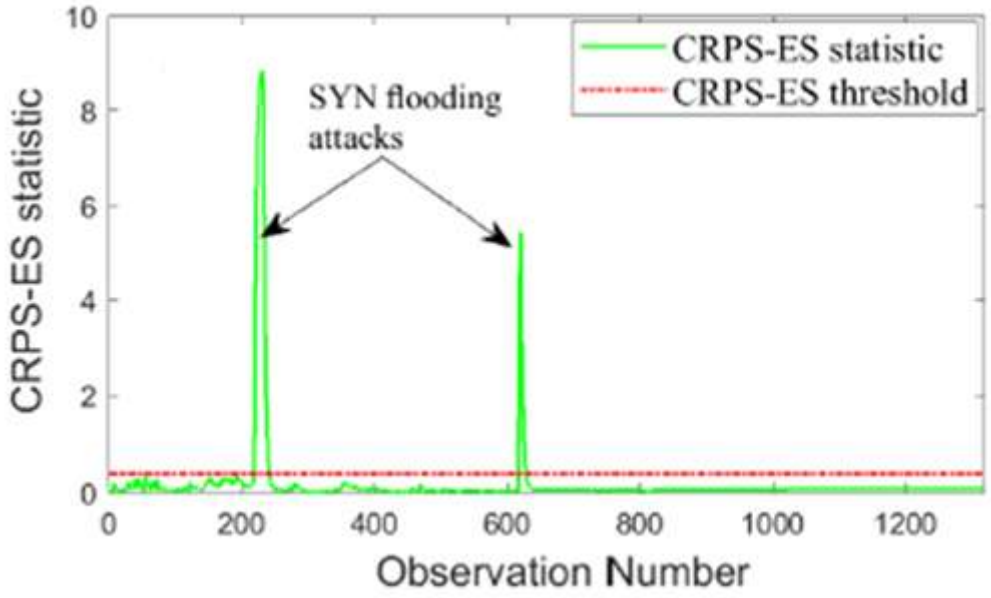

Figure 4.13: Résultat de détection des attaques TCP SYN flood dans W5D2 par CRPS-ES

Les résultats de détection du mécanisme CRPS-ES lorsqu'il est appliqué aux messages ECHO-REPLY reçus dans le cas des attaques SMURF sont reportés dans les figures 4.14, 4.15, 4.16 et 4.17. Ici, la procédure de détection est basée sur le nombre de messages ICMP ECHO-REPLY capturés au niveau de la victime pour chaque temps d'observation. Ces figures indiquent que l'algorithme proposé est capable d'alerter immédiatement toutes les attaques SMURF lorsqu'elles se sont produites. Sur la figure 4.14, la victime a été inondée avec 51681 messages ICMP ECHO-REPLY (tableau 4.1, attaques 1 et 2). Une telle inondation est révélée avec une valeur CRPS de 1.2 qui est beaucoup plus élevée que hes = 0,015. Dans l'attaque SMURF survenue dans W4D3, la victime ciblée a reçu 4455 messages ICMP ECHO-REPLY (figure 4.15); la statistique CRPS correspondante est égale à 7, ce qui dépasse clairement hes. Dans l'attaque SMURF du trafic W4D5 (figure 4.16), 4453 messages ICMP ECHO-REPLY ont été simultanément envoyés à la victime. Dans ce cas, l'attaque de SMURF est caractérisée par une statistique CRPS de 4. Enfin, deux attaques ont été identifiées dans le trafic W5D1 (figure 4.17) par l'algorithme proposé. La première attaque a ciblé la victime durant l'instance 570 avec 6000 messages ICMP ECHO-REPLY (statistique CRPS = 3), et la deuxième attaque contre la victime à l'instance 1914 avec 2655 messages ICMP ECHO-REPLY (statistique CRPS = 1,2).

Dans les deux scénarios, le seuil de détection a de petites valeurs, (hes = 0,7) dans le cas de TCP SYN flood et (hes = 0,51) pour l'attaque SMURF. En pratique, cela reflète la grande sensibilité du mécanisme CRPS-ES. Avec telles valeurs, de petits, voire très petits,

changements dans le comportement du trafic peuvent être révélés. Il s'agit d'une caractéristique appropriée, qui permet à CRPS-ES de gérer les attaques DOS et DDOS modernes faibles à très faibles intensités (Low Rate DOS DDOS : LR-DOS, LR-DDOS).

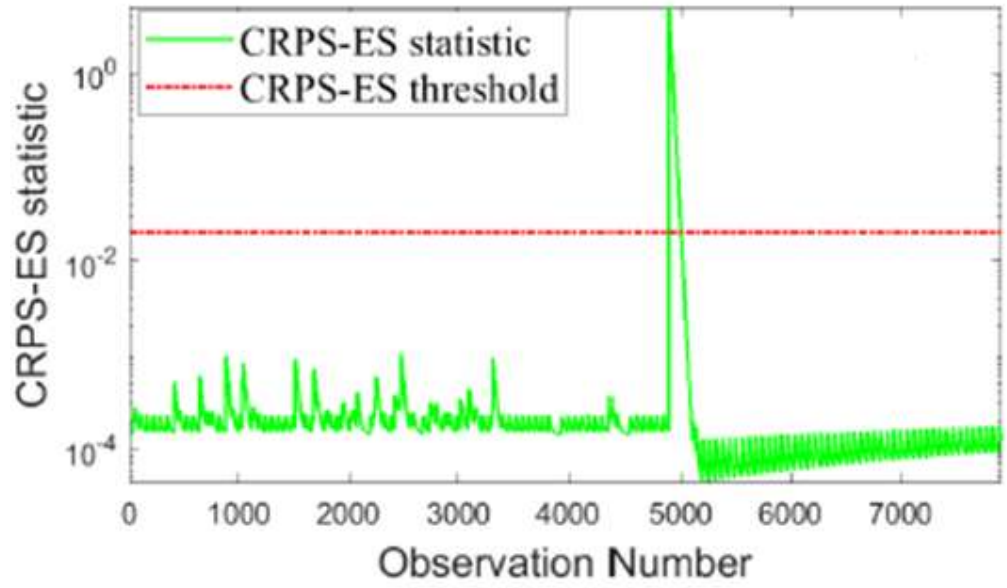

Figure 4.14 : Résultat de détection des attaques SMURF dans W4D1 par CRPS-ES

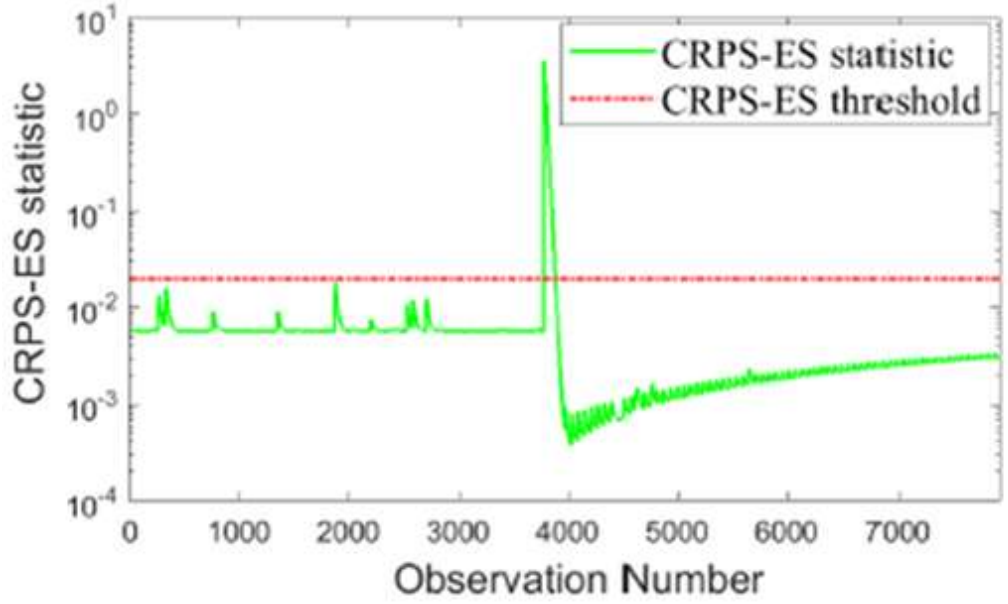

Figure 4.15 : Résultat de détection des attaques SMURF dans W4D3 par CRPS-ES

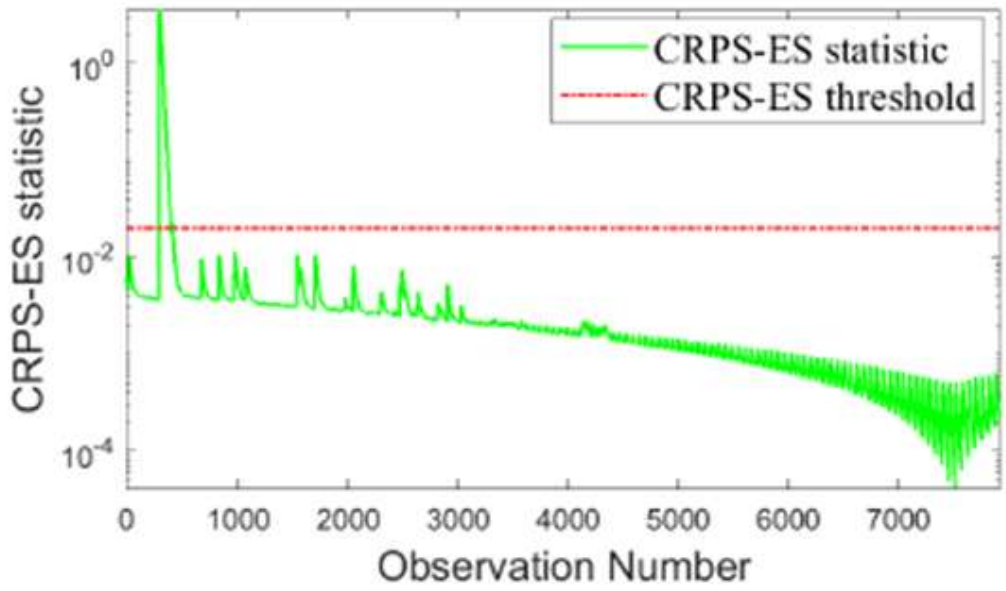

Figure 4.16 : Résultat de détection des attaques SMURF dans W4D5 par CRPS-ES

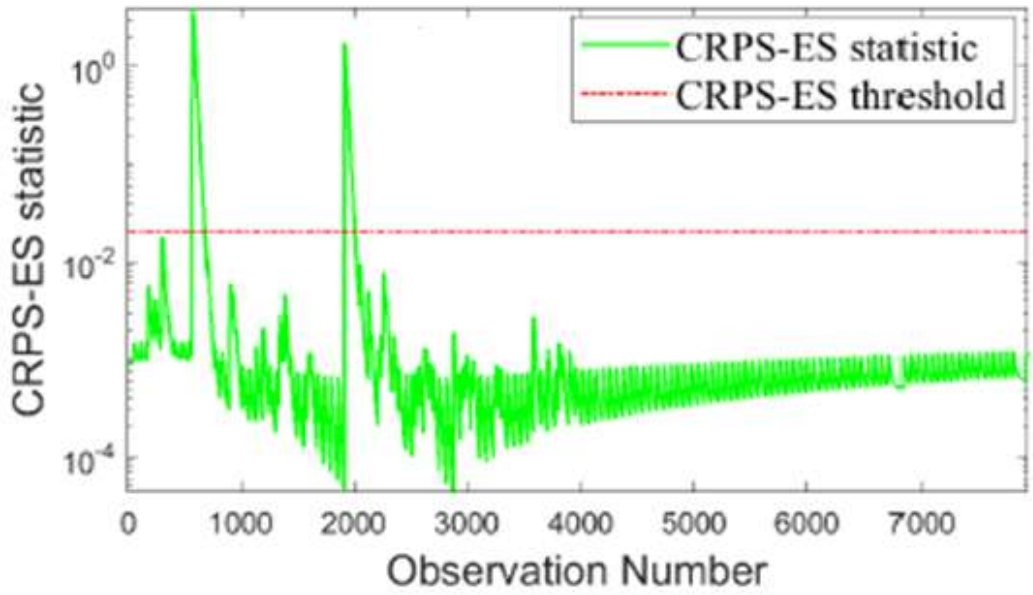

Figure 4.17 : Résultat de détection des attaques SMURF dans W5D1 par CRPS-ES

Pour évaluer quantitativement l'efficacité de détection de la méthode CRPS-ES, les mesures suivantes seront utilisées: TPR, FPR, FNR et AUC. Le tableau 4.2 reporte les performances de détection du mécanisme CRPS-ES lorsqu'il est appliqué à la base DARPA. Les résultats mettent en évidence la capacité de détection du mécanisme CRPS-ES en atteignant un TPR élevé et des FPR et FNR faibles.

Attaque		TPR (%)	FPR (%)	FNR (%)	AUC (%)
TCP SYN flood	Attaque 1	100	0.18	0	99.91
	Attaque 2	97.56	0.15	2	98.70
SMURF	Attaque 1	100	0.45	0	99.77
	Attaque 2	100	0.076	0	99.96
	Attaque 3	100	0.076	0	99.96
	Attaque 4	100	0.076	0	99.96

Tableau 4.2: Performances du CRPS-ES sous la base DARPA99

4.4.2.2. Résultats de détection avec la base MAWI

Ici, nous considérons les scénarios d'attaques par UDP flood et ping flood en utilisant la base MAWI.

La base de données MAWI attire de plus en plus de chercheurs et devient une base de données couramment utilisé dans les études de détection d'intrusion. Toutefois, ce dépôt de données se compose de fichiers TCPDUMP et ne fournissant que le trafic réseau brut sans « ground truth ». Pour évaluer notre approche de détection, nous avons traité la trace de trafic pour séparer le trafic bénin (c.-à-d. exempt d'attaques) et anormal (c.-à-d. avec les attaques) du trafic qui constituent les données d'apprentissage et de test, respectivement. Pour ce faire, tout d'abord, nous avons basé sur le projet d'étiquetage du laboratoire MAWI pour identifier les attaques et les supprimer de la trace de trafic pour construire les données d'apprentissage. Ensuite, les fichiers TCPDUMP bruts sont analysés pour extraire différentes caractéristiques de ces attaques. Le tableau 4.3 présente les caractéristiques des attaques UDP flood et ping flood.

Attaque		Temps d'apparition	Durée
UDP flood	Attaque 1	14:06:50	10s
	Attaque 2	14:11:40	10s
ICMP ping flood	Attaque1	14:05:20	7s
	Attaque 2	14:06:30	11s

Tableau 4.3: Caractéristiques des attaques UDP flood et ping flood dans la base MAWI

La figure 4.18 illustre les résultats de détection de CRPS-ES en présence des attaques UDP flood. Le trafic inspecté contient deux attaques aux instances 41 et 70. Alors que le seuil de détection hes = 0,021, ces deux attaques sont détectées avec précision et leurs mesures CRPS correspondantes sont de 0,037 et 0,044, respectivement. Dans leur étiquetage, les auteurs ont également affirmé la présence de certains comportements suspects, qui sont révélés aux instances 26, 54 et 67. Ces comportements résultent de la forte activité de certains utilisateurs dans l'instance 26, 26% du trafic UDP total est liés à l'utilisateur 163.234.102.228. En effet, le schéma de détection renvoie ces instances un peu plus que le trafic normal.

La figure 4.19 présente les résultats de détection du mécanisme CRPS-ES lorsqu'il est appliqué aux messages ICMP ECHO-REQUEST en présence d'attaques Ping flood. Trois attaques sont détectées aux instances 32, 39 et 48. Ces attaques correspondent à l'analyse du réseau par ICMP, où la plupart des messages ECHO-REQUEST sont liés à trois adresses IP dans 163.234.176.x et deux adresses IP dans 208.108.253.x, représentant 69,5% dans l'instance 32 à 81,7% dans l'instance 48. Ici, hes = 0,009 et les mesures CRPS-ES au cours des trois attaques étaient de 0,011, 0,13 et 0,01 respectivement. Par conséquence, un très petit seuil de détection a été établi. Ici, pour les deux scénarios étudiés, UDP flood et ping flood, les seuils de détection étaient hes = 0,021 et hes = 0,009, respectivement. Ce qui signifie que CRPS-ES peut détecter correctement les attaques à faible intensité lors la considération de telles attaques.

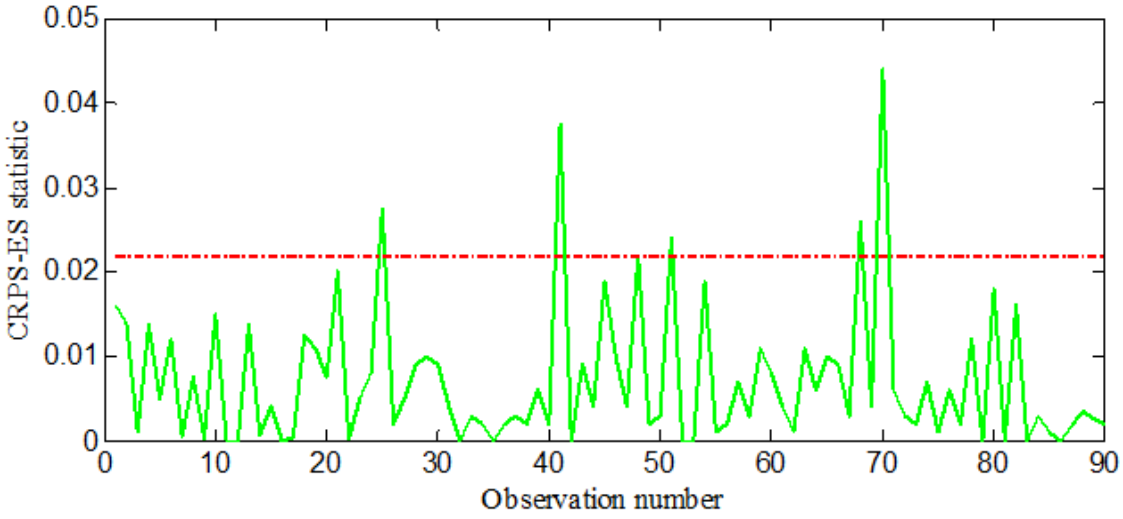

Figure 4.18 : Résultat de détection des attaques UDP flood dans la base MAWI par CRPS-ES

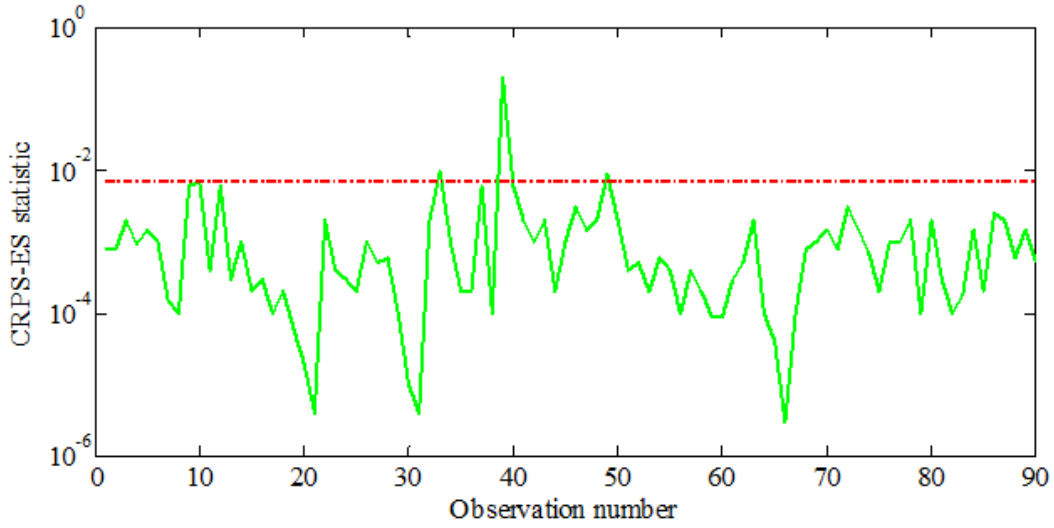

Figure 4.19 : Résultat de détection des attaques ping flood dans la base MAWI par CRPS-ES

Le tableau 4.4 présente l'évaluation des performances du CRPS-ES en termes de TPR, FPR, précision et AUC en fonction des flux anormaux inclus dans la base MAWI.

Attaque		TPR	FPR	Accuracy	AUC
UDP flood	Attaques 1-2	1	0.035	0.967	0.982
Ping flood	Attaque 1	1	0	1	1
	Attaque 2	0.769	0.128	0.857	0.821

Tableau 4.4: Performances de CRPS-ES lors de l'utilisation de la base MAWI

4.4.2.3. Résultats de détection avec la base ICMPv6

Dans cette partie, les performances de CRPS-ES pour révéler les attaques DOS basées sur ICMPv6 sont étudiées. L'ensemble de données se compose d'une collection de trafics ICMPv6 générés sous l'émulateur GNS3 [112]. Ici, nous considérons trois types d'attaques DOS basées sur le protocole ICMPv6 qui sont les attaques NA, NS et RA flood. Leurs détails sont récapitulés dans le tableau 4.5. Nous avons étudié 10mn de trafic anormal, alors que le trafic sans attaque est d'environ 48h. Dans ce cas, le nombre de chaque type de ces messages est contrôlé pour faire face à l'attaque correspondante.

Attaque	Temps d'apparition	Durée
NA flood	1mn20s	4s
NS flood	0s	1s
RA flood	1mn10s	3s

Tableau 4.5: Caractéristiques des attaques DOS à base du protocole ICMPv6

Comme prévu, des valeurs élevées de CRPS qui dépassent le seuil de détection sont obtenues en fonction de la présence d'attaques. Les figures 4.20, 4.21 et 4.22 illustrent les résultats obtenus. Dans ce scénario, les statistiques CRPS-ES atteignent des valeurs élevées lorsque différents types d'attaques se produisent. Dans cet ordre, les (CRPS-ES statistic, hes) correspondants étaient (8, 0,02), (6, 0,018) et (9, 0,019). En effet, les attaques dans ces traces de trafic sont caractérisées par une intensité élevée, ce qui les rend faciles à détecter par le schéma proposé.

Néanmoins, CRPS-ES présentait une sensibilité élevée (c.à.d. de très petites hes) et est adapté pour faire face aux attaques à faible intensité dans cet environnement challengeux.

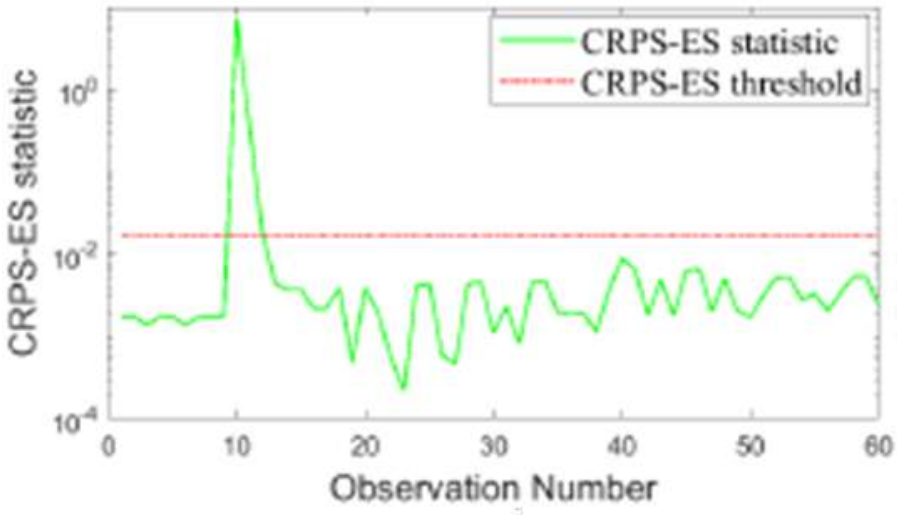

Figure 4.20 : Résultat de détection des attaques NA flood par CRPS-ES

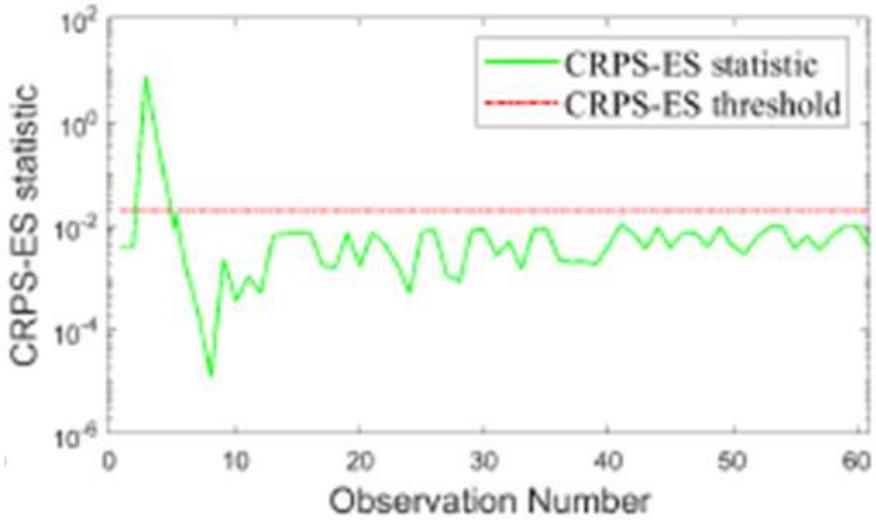

Figure 4.21: Résultat de détection des attaques NS flood par CRPS-ES

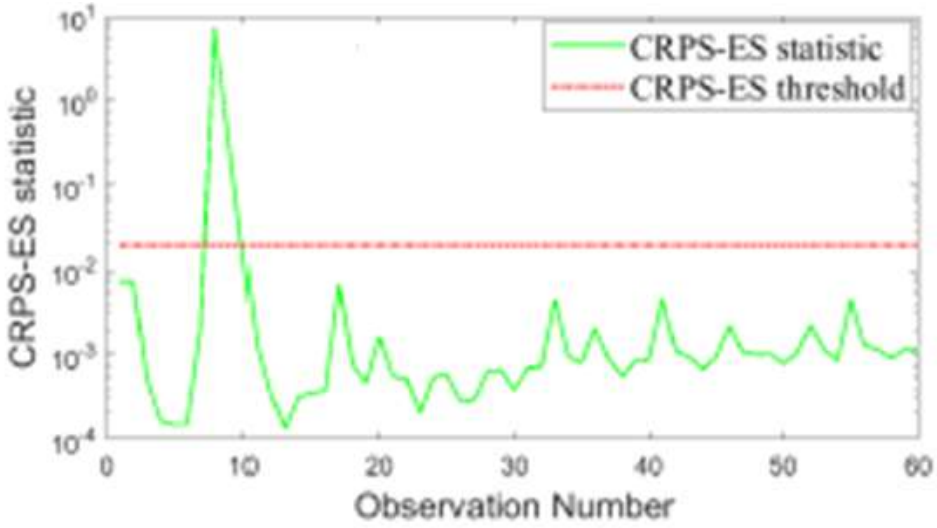

Figure 4.22: Résultat de détection des attaques RA flood par CRPS-ES

Le tableau 4.6 résume les performances de détection du mécanisme CRPS-ES lorsqu'il est appliqué aux traces de trafic ICMPv6. Les résultats confirment que l'approche proposée a de performances prometteuses dans la détection des attaques dans un environnement IPv6.

Attack	TPR	FPR	FNR	AUC
NA flood	1	0	0	1
NS flood	1	0.0327	0	0.984
RA flood	1	0	0	1

Tableau 4.6: Performances de CRPS-ES avec le trafic ICMPv6

4.4.2.4. Comparaison avec des travaux antérieurs

Le tableau 4.7 compare les performances du mécanisme CRPS-ES avec celles de certaines approches, à savoir Anomaly Intrusion Detection (AID) [113], Back Propagation Neural Network (BPN) [114], and Support Vector Machine (SVM) [114] lorsqu'ils sont appliqués au trafic de la base DARPA99. Les résultats montrent que les attaques TCP SYN flood considérées ont été détectées de manière appropriée en utilisant le mécanisme CRPS-ES (TPR = 100% et FPR = 0,18%). Les résultats du tableau 4.7 indiquent que l'approche proposée surpasse ces méthodologies.

Approche	TPR (%)	FPR (%)
CRPS-ES	100	0.18
AID 40x40	96.80	2.85
AID 30x30	96.30	3.15
BPN	96.30	0.70
SVM	99.20	0.84

Tableau 4.7: Comparaison de CRPS-ES avec d'autres approches
(attaques TCP SYN flood de DARPA99)

4.5. Conclusion

Dans ce chapitre nous avons proposé la distance CRPS comme indicateur de détection d'anomalie. L'objectif était de concevoir un schéma efficace pour détecter convenablement les attaques DOS et DDOS. La sensibilité du CRPS aux changements permet de révéler les petites déviations dans le trafic réseau, et de signaler les éventuelles anomalies, y compris les attaques DOS et DDOS en temps réel du fait qu'elle est capable de comparer chaque nouvelle capture de trafic et le trafic de normal de référence.

En premier lieu, nous avons proposé les cartes paramétriques CRPS-Shewhart et CRPS-EWMA. Nous les avons validé pour différents scénarios des attaques TCP SYN flood, dont les résultats obtenus ont montré une amélioration des performances de détection par rapport aux cartes Shewhart et EWMA.

En deuxième lieu, nous avons mis en place un schéma intégré non-paramétrique qui combine la sensibilité de la procédure de lissage exponentiel ES et la bonne capacité du CRPS à séparer les caractéristiques normales et anormales. Ainsi, un seuil automatique non-paramétrique de la statistique CRPS-ES est calculé via la méthode d'estimation de la densité du noyau KDE. Cela offre plus de flexibilité au détecteur CRPS-ES en assouplissant les hypothèses sur la distribution sous-jacente aux données. L'efficacité du CRPS-ES est évaluée à l'aide des bases de trafics réseau DARPA99, MAWI et ICMPv6. Les résultats indiquent que CRPS-ES a montré de bonnes performances par rapport à d'autres algorithmes couramment utilisés.

En suivant la même méthodologie, qu'à CRPS-ES, nous proposerons ainsi dans le chapitre suivant un nouveau mécanisme de détection d'anomalies basé sur la divergence KLD.

Chapitre5

Méthode non-paramétrique de détection d'anomalies à base de la divergence Kullback-Leibler

Dans ce chapitre

- La divergence Kullback-Leibler (KLD)
- KLD dans le domaine de la sécurité
- ES-KLD pour la détection des attaques DOS et DDOS
- Résultats et comparaisons

Chapitre 5

Méthode non-paramétrique de détection d'anomalies à base de la divergence Kullback-Leibler

La détection des anomalies réseau par les mesures de la théorie de l'information est de plus en plus populaire. Ces métriques ont gagné l'intérêt en raison de leur capacité à différencier le trafic légitime du trafic d'attaque avec une faible complexité computationnelle, en utilisant un nombre minimal d'attributs, et peuvent être utilisés à différentes échelles, en termes de nombre d'instances prises par fenêtre de temps. Ces fonctionnalités sont importantes pour détecter les attaques DOS et DDOS.

L'une des mesures les plus courantes est la divergence Kullback-Leibler (KLD), également bien connue en théorie de l'information comme l'entropie relative. KLD est une mesure statistique importante qui peut être utilisée pour quantifier la dissimilarité, la séparation, la distinction ou la proximité entre deux fonctions de densité de probabilité PDF.

Dans ce contexte, la divergence KLD a été largement exploitée dans la détection de nombreux types d'attaques, d'intrusions et de divers problèmes de sécurité qui se produisant dans différentes catégories de réseaux. Cependant, même si plusieurs méthodes de détection qui ont été développées, elles manquent de détection précoce, de précision et conçues au-dessus de larges hypothèses comme la normalité du trafic, seuil de détection manuelle, connaissance préalable de l'attaquant….Pour remédier à ces lacunes, nous proposons dans ce chapitre un mécanisme de détection efficace basé sur KLD, qui permet une détection automatique en temps réel avec une sensibilité élevée sans être limité par la condition de normalité des distributions des trafics contrôlés. Une autre nouveauté supplémentaire dans l'approche proposée, pour améliorer encore les performances de détection, nous avons appliqué un lissage exponentiel ES aux séquences KLD, c'est ES-KLD. Pour le valider, ES-KLD sera utilisé pour la détection des attaques DOS et DDOS, y compris TCP SYN flood, UDP flood, SMURF et les attaques liées au protocole l'ICMPv6, fournies par différentes bases et de traces de trafics réseau.

Le reste du chapitre est organisé comme suit : nous présenterons les différentes approches de détection des cyber-attaques à base de la divergence KLD, en citant leurs principales limitations. Après avoir donné la définition de la KLD, nous introduirons notre technique de détection d'anomalie ES-KLD proposée pour remédier aux limitations susmentionnées. Ensuite, nous détaillerons la méthodologie mise en place pour détecter les attaques DOS et DDOS via ES-KLD. On validera, après, ses performances en considérant plusieurs scénarios d'attaques DOS et DDOS offerts par des bases de trafic réseau IPv4 et IPv6 (DARPA99, MAWI et trafic ICMPv6).

5.1. Applications de la divergence KLD dans le domaine de sécurité

La divergence KLD provient du domaine de la théorie de l'information et est en fait présente à bord de nombreuses disciplines telles que la classification, la reconnaissance de la parole et de l'image [115] [116], le transport [117], l'industrie [118] et la médecine [119]. En outre, KLD a été largement exploité dans le domaine de la sécurité informatique et à la détection de nombreux types d'attaques dans différentes catégories de réseaux.

Li et al. [120] ont présenté une technique de détection des attaques par injection dans des systèmes cyber-physiques utilisant la divergence Kullback-Liebler. Dans [121] Zhang et al. ont appliqué KLD pour superviser les systèmes CPS et détecter les attaques Stealthy Deception.

La divergence KLD a été, déjà, étudiée dans le contexte de la détection de mascarade comme moyen de séparation des données masquées (c.-à-d. les données destinées à imiter l'utilisateur légitime) des données d'attaques. S'ils sont correctement séparés, des scores très élevés peuvent être obtenus. Tapiador et al. [122] ont proposé de détecter les attaques de mascarade en s'appuyant sur une technique qui compare une demande donnée à une demande normale connue à l'aide de la mesure KLD.

Bigi [123] a suggéré KLD pour identifier les auteurs des documents. L'approche construit d'abord un modèle de chaque auteur de document en agrégeant les documents générés par cet auteur. Il développe d'abord un ensemble de modèles de candidats. Ensuite, pour un document donné d'auteur inconnu, l'approche trouve le plus petit KLD entre un modèle connu et le document. Le modèle le plus proche du document est sélectionné comme auteur.

Li et al. [124] ont appliqué le KLD différentiel pour détecter les valeurs anormale des données dans les réseaux de capteurs sans fil.

Dans [52], KLD est proposée pour identifier un ensemble d'applications de logiciels malveillants en se basant sur des SMS.

Dans [125], la divergence KLD des histogrammes des sources et des ports de destination est utilisée pour détecter les tentatives de propagation des vers dans un réseau.

Toutefois, les majorité des techniques de détection utilisant la métrique KLD sont basées sur l'hypothèse de normalité de la distribution du trafic et des données traités. En effet, ces données ont souvent de distributions non normales, ce qui dégrade fortement leur performances. En outre, les seuils de décision dans les techniques conventionnelles basées sur KLD sont généralement prédéfinis manuellement ou ne sont pas utilisés du tout dans le processus de détection. Par conséquence, pour mettre en place une technique de détection d'anomalies à base de la divergence KLD qui répond le mieux aux contraintes imposées par les caractéristiques du trafic réseau, nous proposons ES-KLD avec les extension suivantes :

- Estimation non-paramétrique des distributions de probabilité pour s'adapter aux processus non Gaussiens.
- Lissage exponentiel des séquences KLD, avec implémentation adaptée aux détection en temps réel.
- Etablissement d'un seuil de décison non-paramétrique, automatique et adaptable avec les changements des données utilisés.

Pour vérifier l'efficacité de la technique proposée, ES-KLD sera utilisé pour détecter les attaques TCP SYN flood, UDP flood, les attaques SMURF et les attaques DOS basées sur les messages ICMPv6.

5.2. La divergence KLD

La divergence KLD calcule la différence entre deux distributions de probabilité. Le Kullback-Leibler information entre deux fonctions de densité de probabilité $p_1(x)$ et $p_2(x)$ est définit par [118]:

$$I(p_1 : p_2) = \int_{\mathbb{R}^{d_x}} p_1(x) \log \left[\frac{p_1(x)}{p_2(x)}\right] dx, \qquad (5.1)$$

Et entre $p_2(x)$ et $p_1(x)$ est donnée par :

$$I(p_2 : p_1) = \int_{\mathbb{R}^{d_x}} p_2(x) \log \left[\frac{p_2(x)}{p_1(x)} \right] dx, \qquad (5.2)$$

KLI est non symétrique (c à d $I(p_1 : p_2) \neq I(p_2 : p_1)$) et non négative (c à d $I(p_1 : p_2) \geq 0$ et $I(p_2 : p_1) \geq 0$)).

La distance KLD représente la forme symétrique de KLI [118].

Pour deux distributions $p_1(x)$ et $p_2(x)$, la métrique KLD s'exprime comme suivant [118]:

$$KLD(p_1 : p_2) = I(p_1 : p_2) + I(p_2 : p_1). \qquad (5.3)$$

Pour deux distributions Gaussiennes $p_1 \sim \mathcal{N}(\mu_0, \sigma_0)$ and $p_2 \sim \mathcal{N}(\mu_1, \sigma_1)$, caractérisées respectivement par leurs moyens μ_0 et μ_1 et variances σ_0^2 et σ_1^2, la distance KLD a l'expression analytique suivante [126] :

$$KLD(p_1 \setminus\setminus p_2) = \frac{1}{\sigma_0\sqrt{2\pi}} \int exp\left(\frac{(x-\mu_0)^2}{2\sigma_0^2} \right) \left[\log \frac{\sigma_1}{\sigma_0} - \frac{(x-\mu_0)^2}{2\sigma_0^2} + \frac{(x-\mu_1)^2}{2\sigma_1^2} \right] dx$$

$$= \frac{(\mu_1 - \mu_0)^2}{2\sigma_1^2} + \frac{1}{2}\left(\log \frac{\sigma_1^2}{\sigma_0^2} + \frac{\sigma_0^2}{\sigma_1^2} - 1 \right) \qquad (5.4)$$

Les mesures KLD seront négligeables si $p_1(x)$ et $p_2(x)$ sont trop proches et deviennent importantes lorsque les deux distributions ne sont pas similaires. Par conséquence, KLD peut être utilisé pour révéler des anomalies dans le trafic réseau y compris les attaques DOS et DDOS. Plus précisément, durant le fonctionnement normal du réseau (absence des attaques et des anomalies), de petites valeurs de KLD sont prévues. Sinon, des valeurs élevées apparaissent si le réseau est attaqué ou en présence de d'autres formes anomalies.

5.3. L'approche ES-KLD pour la détection des attaques DOS et DDOS

Nous proposons ici un mécanisme intégré de détection d'anomalies ES-KLD pour détecter les attaques DOS et DDOS par inondation. La méthode proposée est basée sur le lissage exponentiel des mesures KLD.

L'idée de base de cette approche est de calculer d'abord la métrique KLD entre les PDF du trafic capturé et le trafic sans attaque (d'apprentissage). Intuitivement parlant, dans les cas de trafic sans attaques, KLD est proche de zéro et en présence d'attaques, KLD s'écarte considérablement de zéro indiquant la présence d'un événement anormal. Dans le mécanisme

proposé, les séquences KLD sont exponentiellement lissées pour améliorer encore sa sensibilité aux événements anormaux (attaques). La principale raison pour laquelle les mesures de KLD (ES-KLD) sont exponentiellement lissées (ES-KLD) est d'inclure toutes les informations provenant des différents échantillons du trafic.

Définissons la séquence des mesures KLD telle qu'elle est calculée dans (5.3) : KLD $= [d_1 \cdots d_n]$.

La statistique ES-KLD est calculée comme [19]:

$$z_t^{KLD} = \nu d_t + (1-\nu) z_{t-1}^{KLD}, \qquad (5.5)$$

Avec la valeur initiale z_0^{KLD} est la moyenne du vecteur KLD μ_0^{KLD} dans le cas sans attaques.

A partir de (5.5), nous notons que la statistique de décision ES-KLD intègre l'information des observations passées et actuelles dans le processus de décision, ce qui contribue à améliorer sa sensibilité aux petits changements. Pour démontrer explicitement ce point, l'ES-KLD est exprimé de façon récursive comme [19]:

$$z_t^{KLD} = \nu d_t + (1-\nu)\overbrace{[\nu d_{t-1} + (1-\nu) z_{t-2}^{KLD}]}^{z_{t-1}^{KLD}}$$

$$= \nu d_t + \nu(1-\nu) d_{t-1} + (1-\nu)^2 z_{t-2}^{KLD} \qquad (5.6)$$

En utilisant l'équation (5.6) récursivement, nous obtenons :

$$z_n^{KLD} = \nu d_n + \nu(1-\nu) d_{n-1} + \nu(1-\nu)^2 d_{n-2} + \cdots$$

$$+\nu(1-\nu)^{n-1} d_1 + \nu(1-\nu)^n d_0 \qquad (5.7)$$

L'équation (5.7) peut aussi être exprimée en forme compacte comme :

$$z_t^{KLD} = \nu \sum_{t=1}^{n} (1-\nu)^{n-t} d_t + (1-\nu)^n d_0, \quad (5.8)$$

où $\nu(1-\nu)^{n-t}$ représente le poids du d_t, qui diminue de façon exponentielle pour les échantillons passés. Essentiellement, la valeur de ν est pratiquement sélectionnée entre 0,2 et 0,3 pour distinguer les petites déviations.

Nous définissons un seuil de détection non-paramétrique comme le (1-α) ième quantile de la distribution estimée de la statistique ES-KLD, z^{KLD}, à l'aide de l'estimation de la densité du noyau (KDE). Une anomalie est détectée lorsque la fonction de décision ES-KLD dépasse le seuil de détection. La procédure générale de détection des attaques DOS et DDOS par l'ES-KLD est illustré à la figure 5.1.

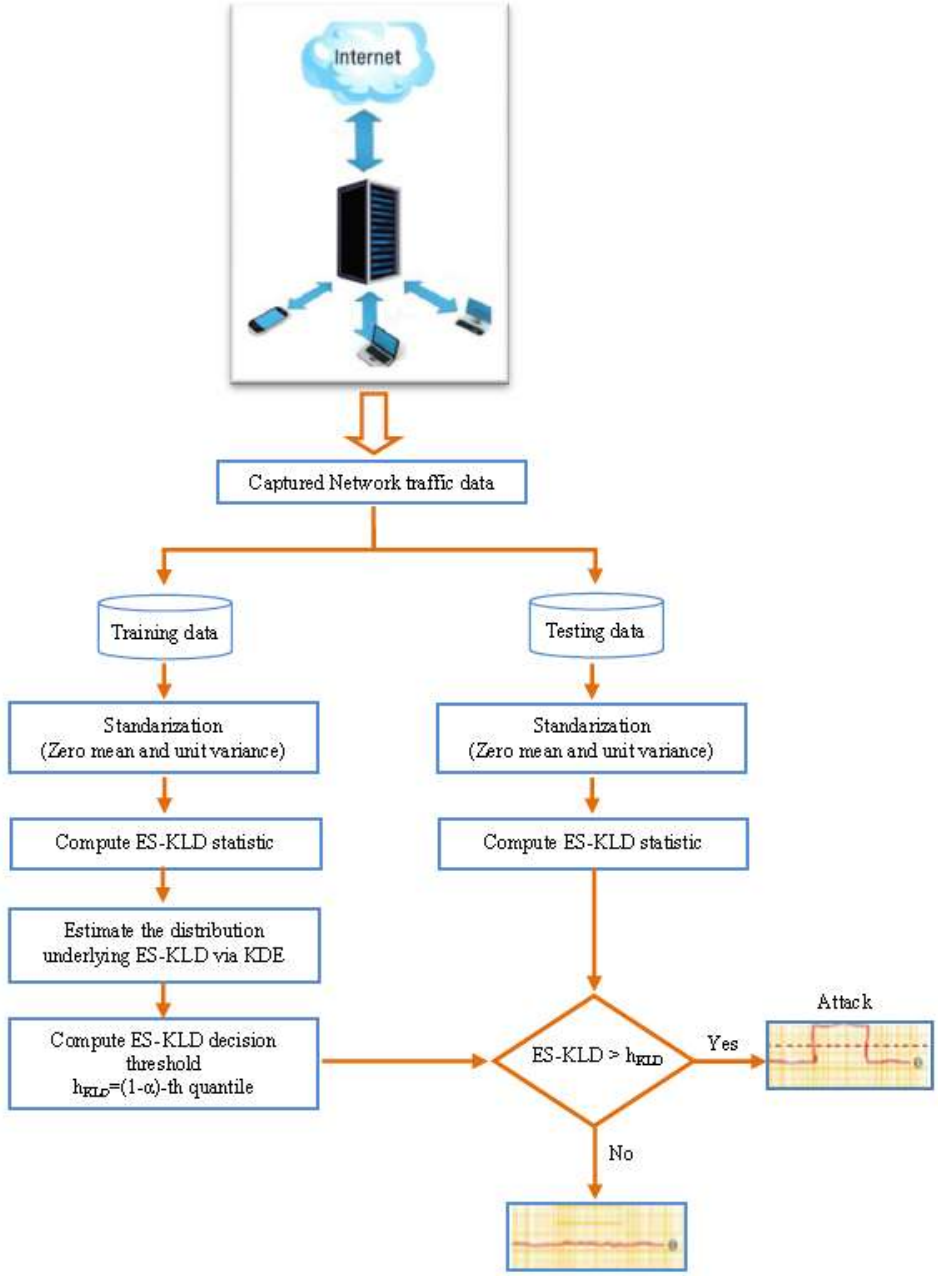

Figure 5.1: La procédure générale de détection des attaques DOS et DDOS par ES-KLD

5.4. Validation et résultats de détection

Nous examinons la capacité du mécanisme proposé ES-KLD à identifier certains types d'attaques DOS et DDOS. Précisément, les attaques TCP SYN flood, UDP flood, SMURF et les attaques DOS à base ICMPv6 seront étudiées. A cette fin, de nombreux scénarios sont menés en utilisant le trafic réseau IP à partir de trois bases de données DARPA99 (cf. paragraphe 3.4), MAWI et ICMPv6 (cf. paragraphe 4.3).

Ici, nous évaluons les performances de l'approche ES-KLD en considérant les caractéristiques de différents protocoles :

- Les attaques TCP SYN flood sont contrôlées en fonction du flux des segments SYN et du flux total TCP reçus par temps d'échantillonnage.
- Les attaques UDP flood sont inspectées à l'aide du nombre de datagrammes UDP et de messages destination inaccessible de l'ICMP.
- Les attaques ping flood sont détectées à l'aide des messages ICMP ECHO-REQUEST reçus et de l'ensemble du trafic ICMPv4.
- Les attaques SMURF sont identifiées en fonction du nombre des messages ICMP ECHO-REPLY reçus et de l'ensemble du trafic ICMPv4.
- Les attaques DOS basées sur le protocole ICMPv6 sont détectées en vérifiant les messages NS, NA et RA. Dans ce cas, nous contrôlons le flux de ces messages ainsi que l'ensemble du trafic ICMPv6.

Pour extraire les caractéristiques pertinentes (ex., segments TCP SYN et flux TCP….) des bases de données du trafic réseau, un prétraitement des trois bases de données est effectué (cf. paragraphe 3.4).

5.4.1. Résultats de détection avec DARPA99

La base DARPA99 fournit différents types d'attaques, y compris TCP SYN flood, UDP flood et les attaques SMURF. Les détailles de ces attaques sont récapitulées dans le tableau 4.1

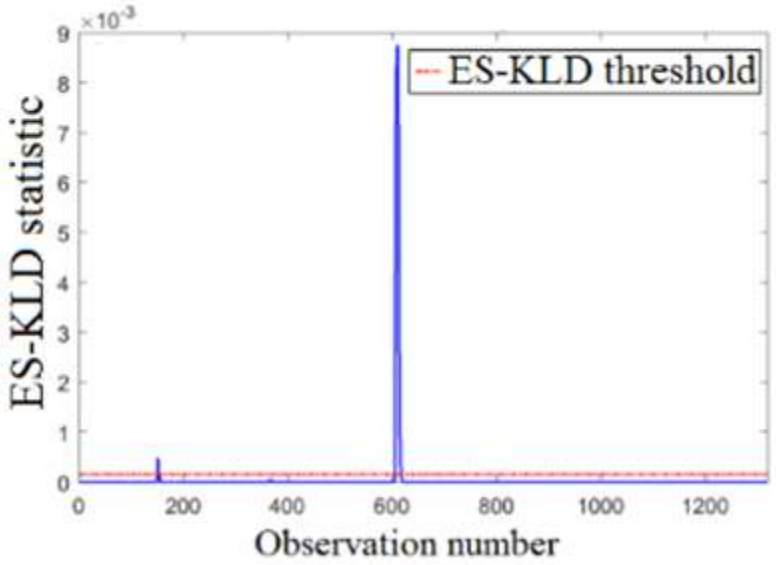

Figure 5.2: Résultat de détection en présence des attaques TCP SYN flood (W5D1, flux de segments SYN)

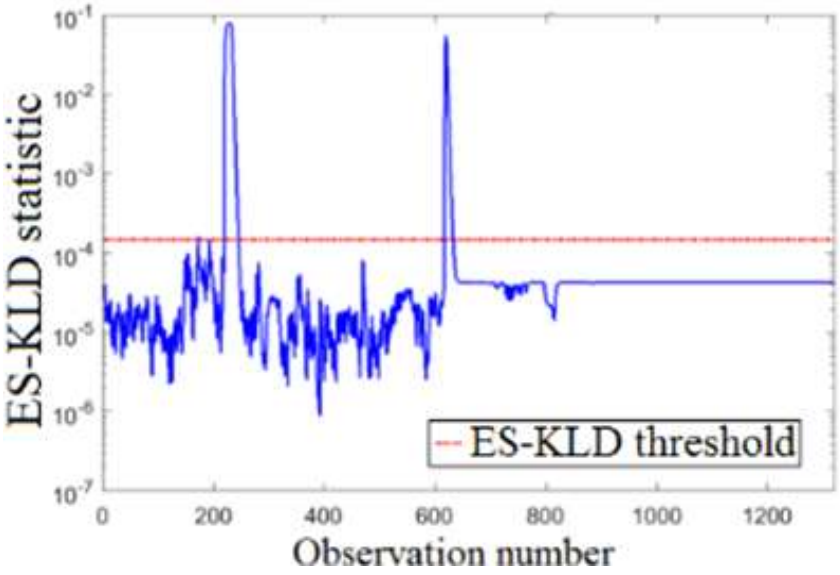

Figure 5.3 : Résultat de détection en présence des attaques TCP SYN flood (W5D2, flux de segments SYN)

Les figures 5.2 et 5.3 montrent les résultats du mécanisme ES-KLD en fonction des segments SYN reçus. Tout d'abord, ces résultats confirment la sensibilité élevée de l'ES-KLD aux changements. Par conséquence, de petites valeurs des statistiques ES-KLD (0 ou près de 0) sont obtenues lorsque le trafic surveillé est exempt d'attaques. Sinon, avec l'apparition d'attaques TCP SYN flood, le trafic anormal résultant devient trop différent du trafic normal sans attaque, de sorte que les statistiques ES-KLD augmentent de manière significative et de

grandes valeurs sont enregistrées. En outre, de petits seuils de détection sont établis, ce qui signifie que ES-KLD peut révéler même les attaques DOS et DDOS de faible intensité.

Dans la figure 5.2 qui représente la détection basée sur le trafic W5D1, l'attaque s'est produite aux instances 605 à 611. Au cours de cette attaque, la victime (@IP=172.16.112.50) a reçu en moyenne 2928 segments SYN par temps d'observation.

La figure 5.3 montre la présence de deux attaques TCP SYN flood dans le trafic W5D2. Lors de la première attaque, la victime a été submergée par une moyenne de 3027 SYN/temps d'observation entre les instances 219 et 232. Dans la deuxième attaque, la victime 192.168.1.1 a été inondée par 10256 segments SYN entre les instances 617 et 620.

Les figures 5.2 et 5.3 indiquent que les attaques TCP SYN flood sont détectées avec succès et que le taux de détection du mécanisme ES-KLD a atteint 100 %. Dans les deux scénarios, il est clairement illustré que les statistiques ES-KLD augmentent proportionnellement avec l'intensité de l'attaque DOS.

En outre, la figure 5.2 indique que le mécanisme ES-KLD révèle deux autres événements anormaux aux instances 150 et 151. En effet, au cours de ces deux événements, différents clients ont demandé de nombreuses connexions TCP, à l'exception du propriétaire de l'adresse IP 202.77.162.213 qui a envoyé au serveur 172.16.114.50 (hébergeur de serveurs Web) 1007 et 561 segments SYN aux instances 150 et 151, respectivement. Selon les auteurs de la base de données DARPA99, 20 segments SYN envoyés simultanément peuvent paralyser la victime. Par conséquence, le trafic dans ces deux cas a certainement un comportement anormal et peuvent être considéré donc comme des événements anormaux. Selon la documentation de la base de données, on constate que procédure d'attaque consiste à ce que l'attaquant envoie 20 segments SYN à chaque port de la victime. Après avoir analysé le fichier TCPDUMP W5D1 brut, on remarque que la source ne suit pas cette procédure. Le client 202.77.162.213 a envoyé tous les segments SYN au port 80 du 172.16.114.50. Nous pouvons conclure, donc, qu'il ne s'agit pas d'une attaque. Mais, il reste un trafic anormal avec un pic de 39 segments SYN/s.

Ainsi, à la figure 5.3, un pic à l'instance 171 dépasse le seuil de décision établit. Dans ce scénario, la génération de segments SYN ne suit pas la procédure d'attaque adoptée par les auteurs. Plus précisément, l'utilisateur ayant l'adresse IP 172.16.112.207 a envoyé en 1 seconde environ 27 segments SYN au port 80 de l'adresse IP 206.246.131.226 et le client utilisant l'adresse IP 206.246.131.226 et le client utilisant l'adresse IP 172.16.115.5 envoyaient en 4 seconds 73 et 48 segments SYN au port 80 du 207.18.199.3 et 209.67.29.11,

respectivement. Ici, on peut noter que (i) il y a quelques erreurs dans l'étiquetage des attaques (par exemples, les attaques UDP flood dans W5D1 et SMURF dans W4D5). ii) Il y a des critiques au sujet du trafic normal qui peut fortement affecter les données d'apprentissage; le réseau présente une activité très faible, semble même dans certaines traces à court de service.

Maintenant, les performances du mécanisme ES-KLD dans la détection des attaques TCP SYN flood sont étudiées lorsque l'ensemble du flux TCP est surveillé.

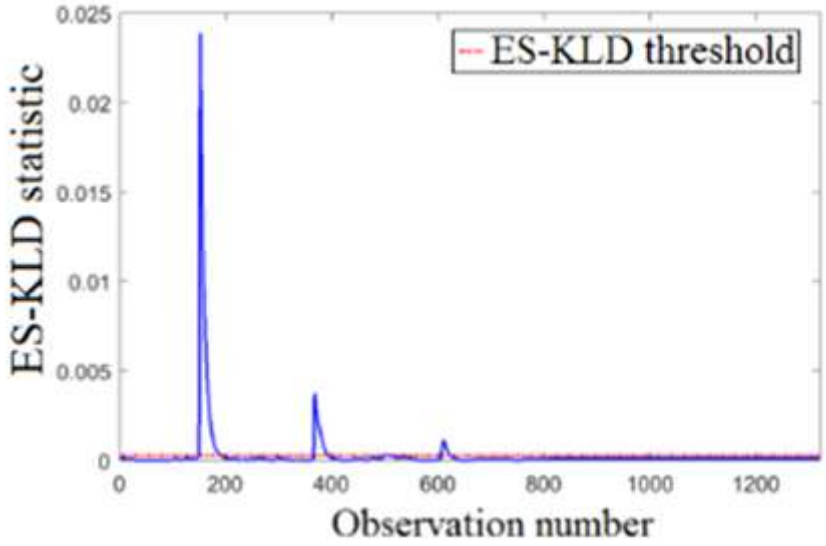

Figure 5.4 : Résultat de détection en présence des attaques TCP SYN flood (W5D1, flux TCP)

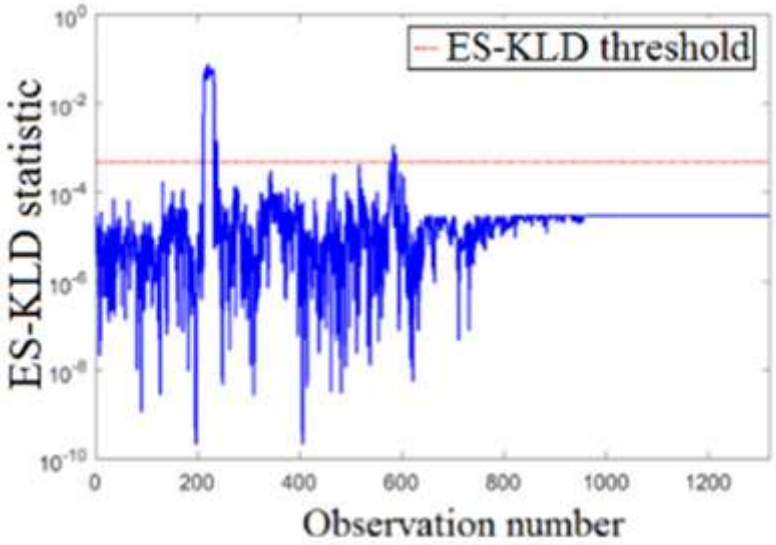

Figure 5.5 : Résultat de détection en présence des attaques TCP SYN flood (W5D2, flux TCP)

Les figures 5.4 et 5.5 affichent les résultats de détection du mécanisme ES-KLD lors du contrôle des flux TCP. Bien que les attaques soient détectées, on peut remarquer que la surveillance de tous les flux TCP ne peut pas conduire nécessairement à une bonne précision de détection, il ne s'agit pas donc, d'un bon indicateur des attaques TCP SYN flood. En tant que protocole connecté, différents segments (e.g. Acknowledgment, Push, and Reset) sont utilisés pour maintenir la qualité des sessions établies. En effet, ces segments peuvent avoir une forte contribution dans l'ensemble du flux TCP et peuvent donc fortement affecter la détection de l'attaque TCP SYN flood. Dans la figure 5.4, les segments Acknowledgment, Push, and Reset représentent respectivement 84,2 %, 93,12 % et 95,07 % du flux TCP global aux instances 150, 366 et 367. Ils peuvent même masquer les attaques dans certaines situations, en particulier dans les événements flash crowd et la mauvaise qualité de lien dans lesquels plus de segments Acknowledgment et les demandes de retransmission seront transmis. L'attaque SYN est à peine détectée; le nombre de segments SYN représente moins de 35 % du flux TCP dans l'instance 605 et environ 41 % aux instances 606 et 607.

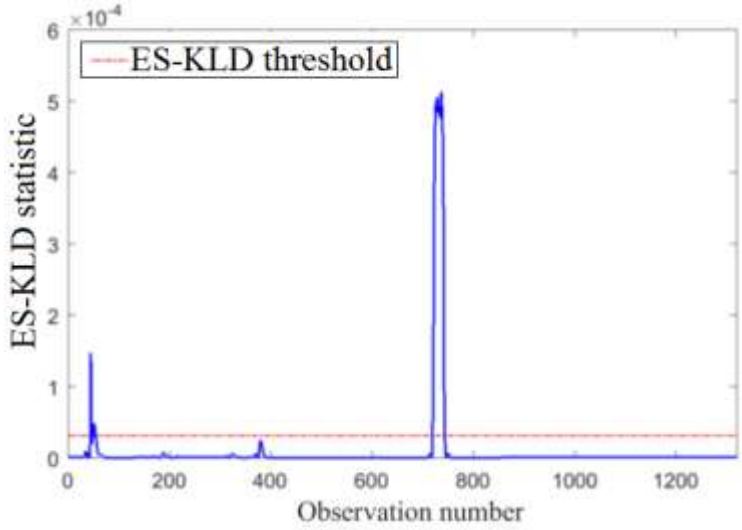

Figure 5.6 : Résultat de détection en présence des attaques UDP flood (W5D1)

La figure 5.6 illustre les résultats de détection du mécanisme ES-KLD en présence des attaques UDP flood. La figure indique que ces attaques ont été détectées avec succès. Au cours de ces attaques, les victimes (@IP 172.16.112.50 et 172.16.113.50) ont échangé entre elles 272742 UDP datagrammes. Plus précisément, chacune a envoyé 136371 datagrammes UDP au port Echo de l'autre. Nous pouvons observer que, pendant les deux attaques, les statistiques ES-KLD dépassent largement le seuil de détection. A la figure 5.6, nous observons également la présence de certains pics qui ont commencé à l'instance 45. Au cours

de ces instances, l'utilisateur ayant l'adresse IP 172.16.112.100 a continué d'envoyer des requêtes DNS (un total de 12472 requêtes) même si le serveur DNS (@IP 172.16.112.20) indique que les noms demandés n'existent pas.

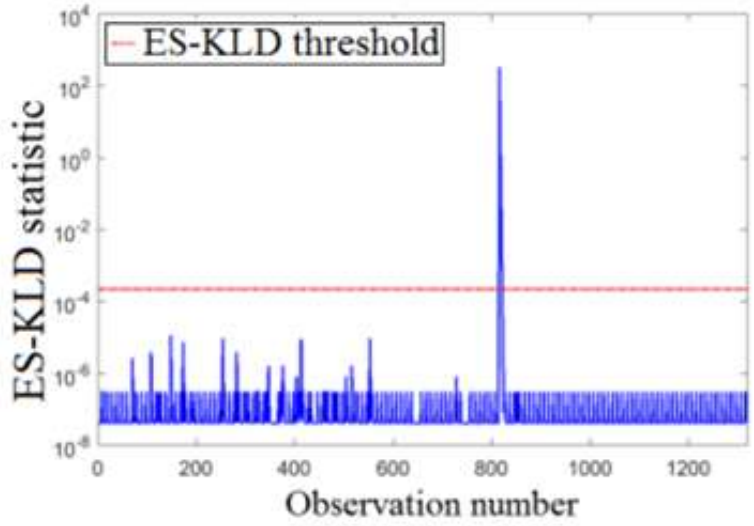

Figure 5.7 : Résultat de détection en présence des attaques SMURF (W4d1, messages ICMP ECHO-REPLY)

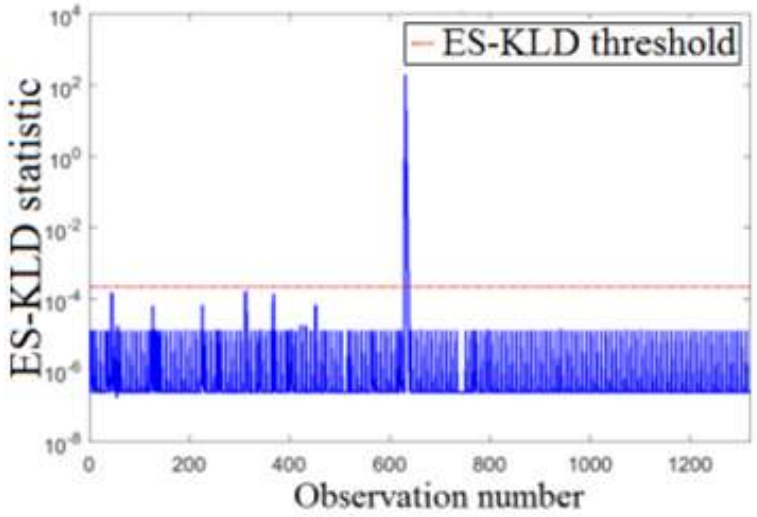

Figure 5.8 : Résultat de détection en présence des attaques SMURF (W4d3, messages ICMP ECHO-REPLY)

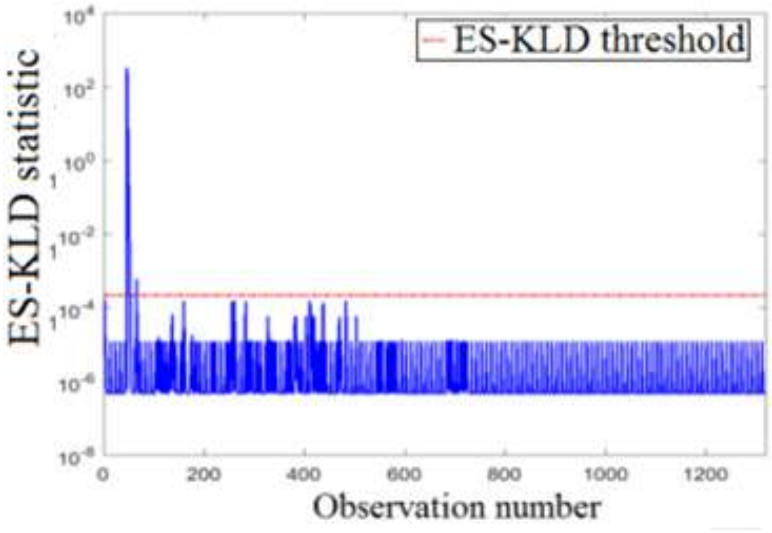

Figure 5.9 : Résultat de détection en présence des attaques SMURF (W4d5, messages ICMP ECHO-REPLY)

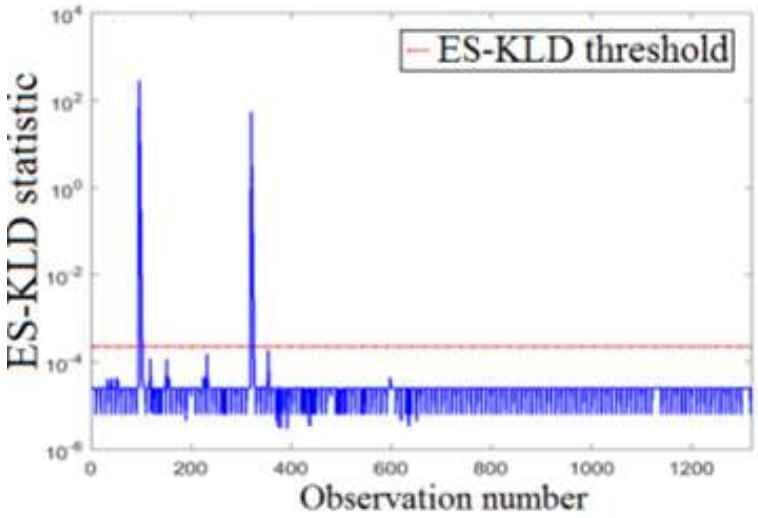

Figure 5.10 : Résultat de détection en présence des attaques SMURF (W5d1, messages ICMP ECHO-REPLY)

Les figures 5.7, 5.8, 5.9 et 5.10 représentent les résultats de détection du mécanisme ES-KLD lorsqu'il est appliqué aux messages ICMP ECHO-REPLY reçus durant les attaques SMURF. Toutes les attaques considérées sont signalées par le mécanisme conçu. Dans la figure 5.7, les attaques SMURF se sont produites à l'instance 815, et leur victime (@IP 172.16.112.50) a été inondée avec 51681 messages ECHO-REPLY. La figure 5.8 montre que le trafic W4D3 comprend une attaque du SMURF à l'instance 630, dans laquelle la victime

visée (@IP 172.16.112.100) a reçu 4455 messages ECHO-REPLY. A la figure 5.9, l'attaque SMURF du trafic de W4D5 s'est produite à l'instance 46, où 4453 messages ECHO-REPLY ont été envoyé à la victime (@IP 172.16.112.50). La figure 5.10 révèle deux attaques dans le trafic W5D1. La première attaque a visée @IP 172.16.112.50 à l'instance 95 avec 6000 messages ECHO-REPLY, et la deuxième attaque a été contre @IP 172.16.114.50 à l'instance 319 avec 2655 messages ECHO-REPLY.

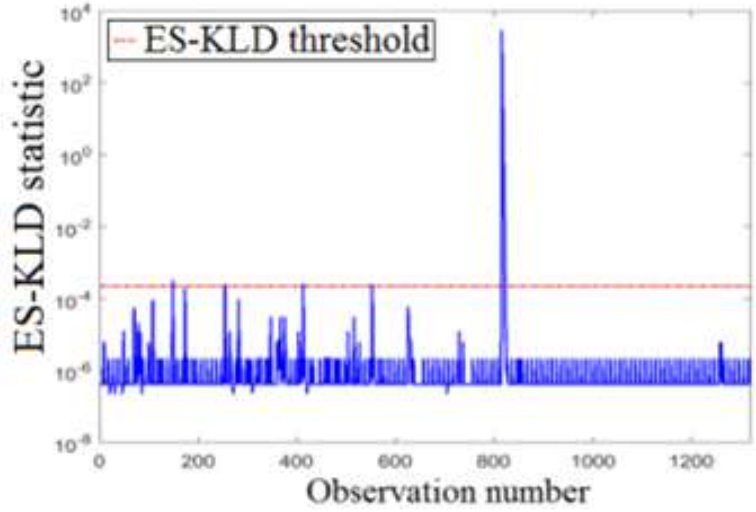

Figure 5.11 : Résultat de détection en présence des attaques SMURF (W4d1, flux ICMP)

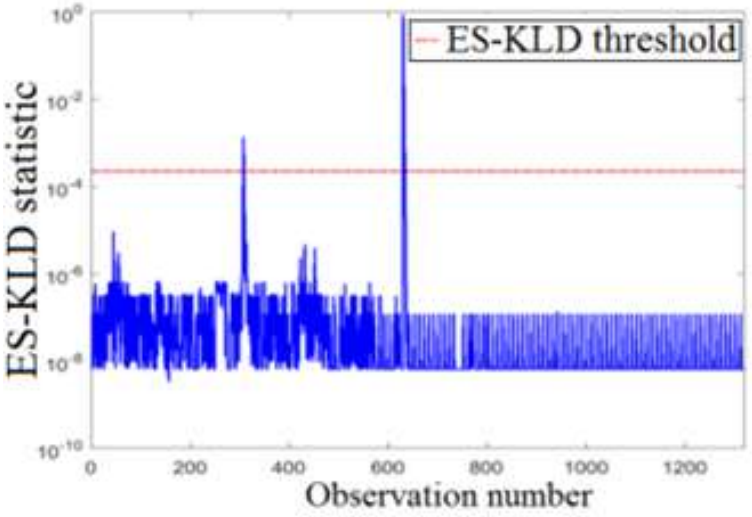

Figure 5.12 : Résultat de détection en présence des attaques SMURF (W4d3, flux ICMP)

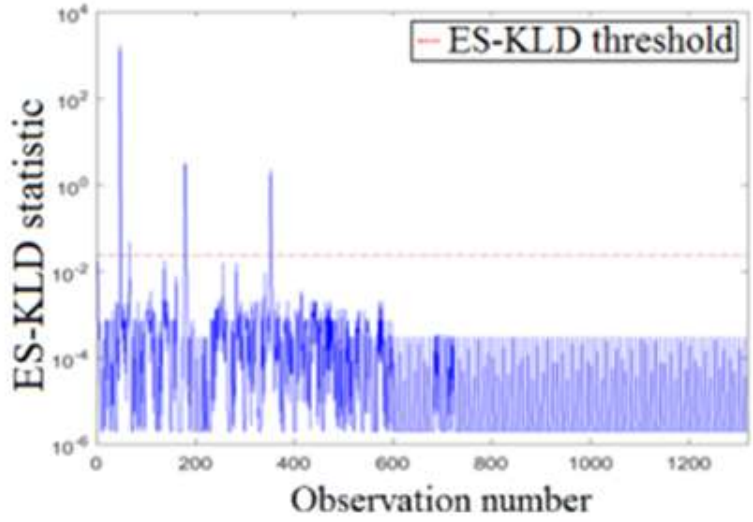

Figure 5.13 : Résultat de détection en présence des attaques SMURF (W4d5, flux ICMP)

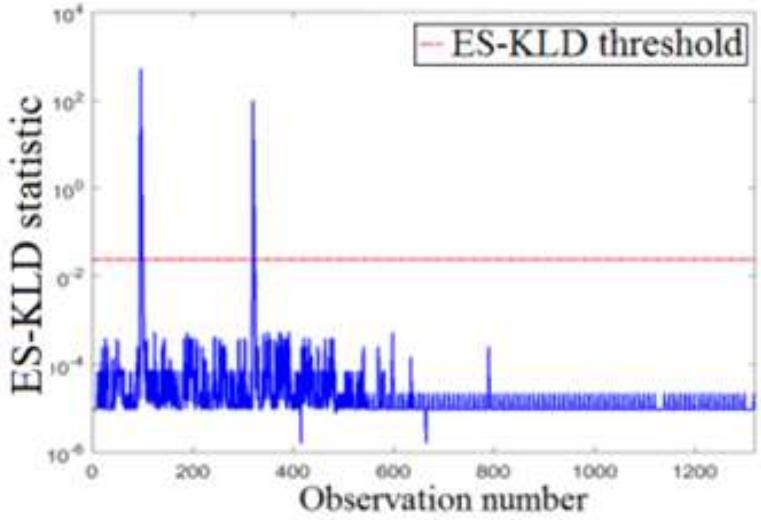

Figure 5.14 : Résultat de détection en présence des attaques SMURF (W5d1, flux ICMP)

Les figures 5.11, 5.12, 5.13 et 5.14 montrent les résultats de détection de l'algorithme ES-KLD lorsqu'il est appliqué au flux ICMP en présence d'attaques SMURF. Ces résultats indiquent que le trafic ICMP peut être utilisé pour détecter les attaques SMURF. Néanmoins, il peut générer de fausses alarmes en raison des autres messages ICMP qui sont incorporés dans le trafic ICMP. Dans les figures 5.12 et 5.13, les messages de destination et port inaccessible sont responsables de ces deux anomalies dans W4D3 et des trois anomalies dans les trafics W4D5 où ils représentent 100 % du trafic ICMP. Dans la figure 5.13, durant les

deux instances 307 et 308, il y a 267 messages destination inaccessible, tandis que pas de message ECHO-REPLY a été généré. De plus, les messages ECHO-REPLY représentent 0 % du trafic ICMP (figure 5.14) aux instances 178, 179 et 352, bien qu'il compte respectivement 192,188 et 167 messages.

Le tableau 5.1 récapitule les performances de détection du mécanisme ES-KLD en termes de TPR, FPR, Accuracy et AUC. On peut remarquer que l'algorithme ES-KLD fournit des performances satisfaisantes dans la détection des attaques DOS de DARPA99 avec de faibles FPR et des TPR élevés. Cela confirme l'efficacité de détection du mécanisme ES-KLD. La dégradation des performances observées dans TCP SYN flood (attaque 2) et UDP flood (attaque1) sont liées aux spécifications de la base DARPA. Sinon, d'autres anomalies (non citées par les auteurs) sont révélées, et par conséquence, des TPR plus élevés (moins de FPR) seront obtenues.

Attaque		TPR	FPR	Accuracy	AUC
TCP SYN flood	Attaque 1	1	0.002	0.998	0.999
	Attaque 2	0.976	0.002	0.998	0.987
UDP flood	Attaque 1	0.917	0.001	0.998	0.958
SMURF	Attaque 1	1	0.005	0.995	0.998
	Attaque 2	1	0.001	0.999	0.998
	Attaque 3	1	0.001	0.999	1
	Attaque 4	1	0.001	0.999	1

Tableau 5.1: Performances de détection de l'ES-KLD avec la base DARA99

5.4.2. Résultats de détection avec la base MAWI

Ici, nous évaluons les performances de notre approche de détection en présence des attaques SYN flood dans la base MAWI. Pour séparer le trafic bénin du trafic anormal, nous reprenons la même procédure utilisée dans le paragraphe 4.4.2.2. Ensuite, les fichiers TCPDUMP bruts sont analysés (même procédure de prétraitement que le paragraphe 3.4) pour extraire différentes caractéristiques de ces attaques. Le tableau 5.2 présente les caractéristiques des attaques TCP SYN flood.

Attaque		Temps d'apparition	Durée
TCP SYN flood	Attaque 1	14:01:19	2s
	Attaque 2	14:06:01	1s
	Attaque 3	14:06:02	2s
	Attaque 4	14:12:08	5s
	Attaque 5	14:13:33	1s
	Attaque 6	14:13:34	2s
	Attaque 7	14:13:44	2s

Tableau 5.2: Caractéristiques des attaques TCP SYN flood dans la base MAWI

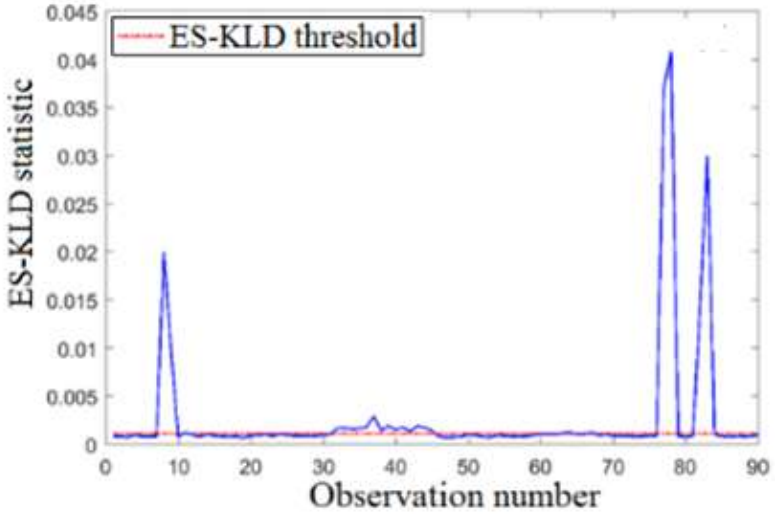

Figure 5.15 : Résultat de détection en présence des attaques TCP SYN flood (Base MAWI, flux de segments SYN)

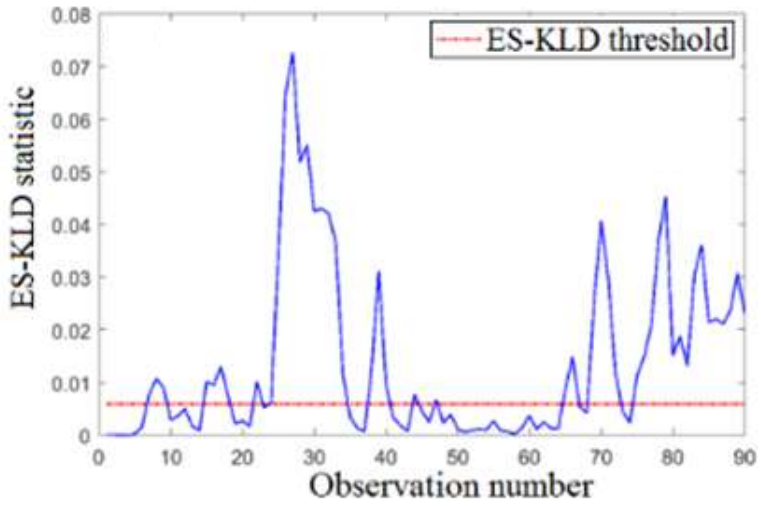

Figure 5.16 : Résultat de détection en présence des attaques TCP SYN flood (Base MAWI, flux TCP)

Les figures 5.15 et 5.16 montrent les résultats de détection de l'ES-KLD en présence d'attaques TCP SYN flood dans la base de données MAWI. A la figure 5.15, les mesures des segments SYN reçus révèlent 7 attaques (aux instances : 8, 9, 37, 77, 78, 82 et 83) correspondant aux attaques de balayage de port SYN. Par exemple, à l'instance 8, l'attaquant (@IP : 89.56.8.149) a scanné environ de 12623 adresses IP du 208.108.134.x au 208.108.251.x. Le tableau 5.3 fournit les détails de ces attaques.

Instance d'attaque	@IP de l'attaquant	Plage @IP scannnées	Nombre des @IP scannées
8	89.56.8.149	208.108.134.x to 208.108.251.x	12623
9	89.56.8.149	208.108.135.x to 208.108.191.x	7875
37	209.8.83.176 89.56.8.149	209.67.86.x to 209.67.91.x 199.199.56.x to 199.199.63.x	1312 1239
77	93.99.192.106	136.172. 128.x to 136.172.255.x	18382
78	93.99.192.106	136.172.0.x to 136.172.235.x	19323
82	200.59.160.64	209.72.0.x to 209.72.159.x	10243
83	200.59.160.64	209.67.64.x to 209.67.127.x	16183

Tableau 5.3: Attaques TCP SYN flood dans la base MAWI

5.4.3. Résultats de détection avec la base de trafic ICMPv6

La base de trafic ICMPv6 comprend une série d'attaques DOS basées sur ICMPv6. Ici, nous étudions la capacité du mécanisme ES-KLD dans la détection des attaques par inondation qui utilisent les messages ICMPv6 de types NA, NS et RA. Les caractéristiques de ces attaques sont reportées dans le tableau 4.4.

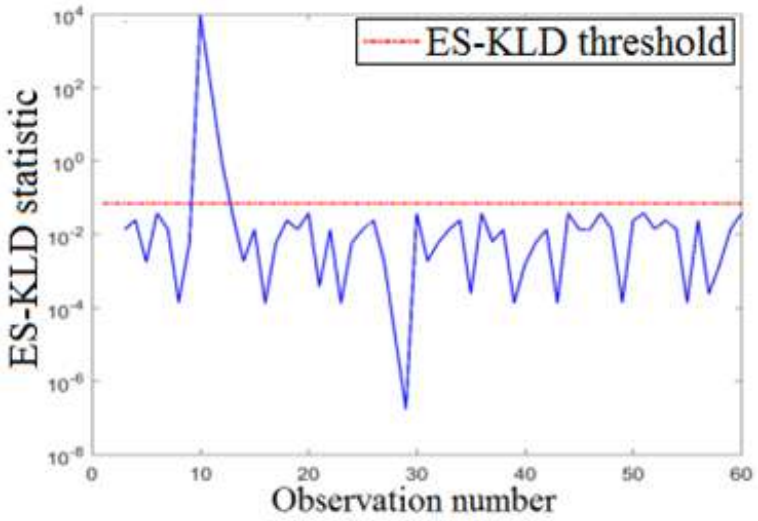

Figure 5.17 : Résultat de détection en présence des attaques NA flood (Flux de messages NA)

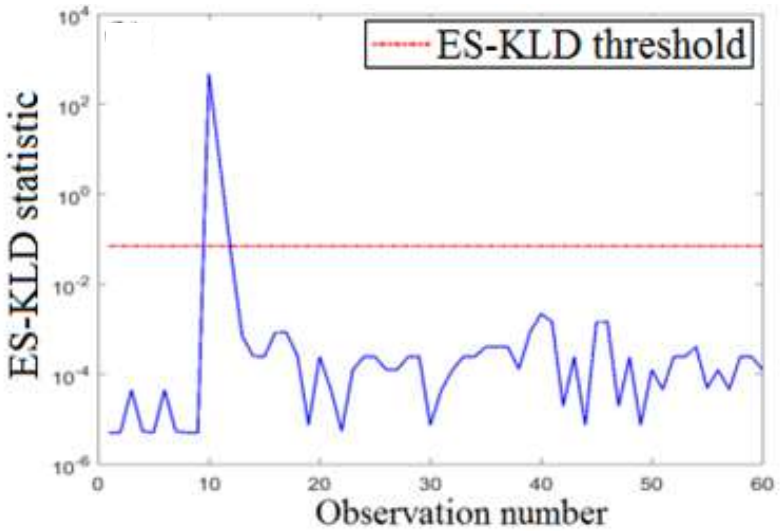

Figure 5.18 : Résultat de détection en présence des attaques NA flood (Trafic ICMPv6)

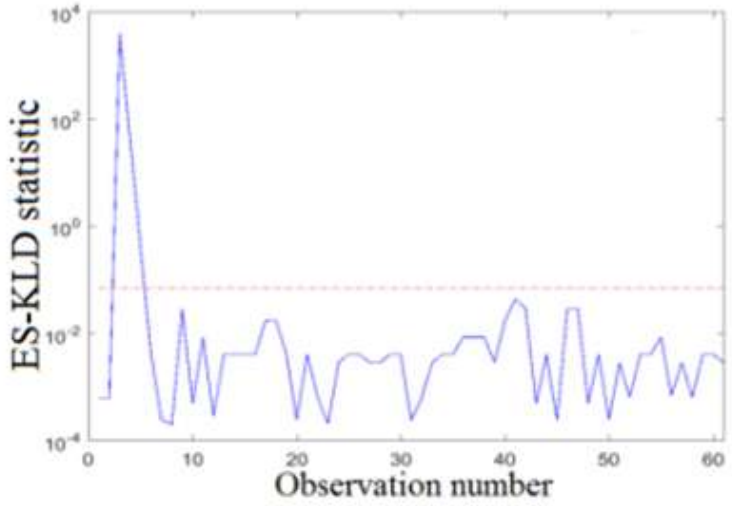

Figure 5.19 : Résultat de détection en présence des attaques NS flood (Flux de messages NS)

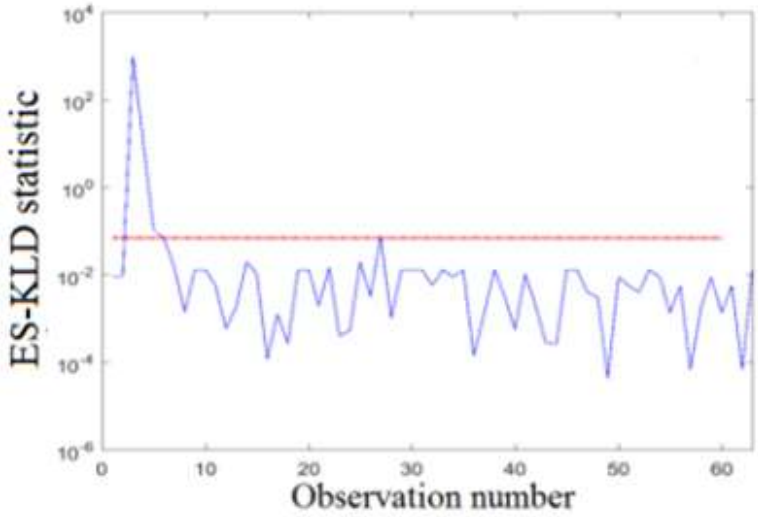

Figure 5.20 : Résultat de détection en présence des attaques NS flood (Trafic ICMPv6)

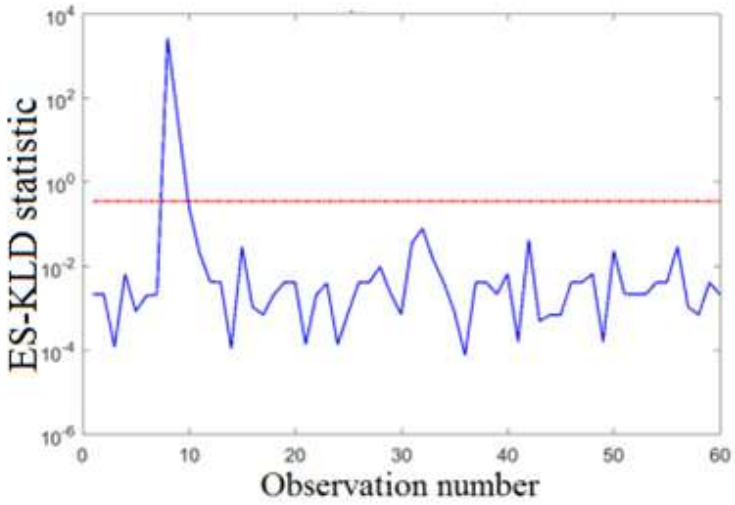

Figure 5.21 : Résultat de détection en présence des attaques RA flood (Flux de messages RA)

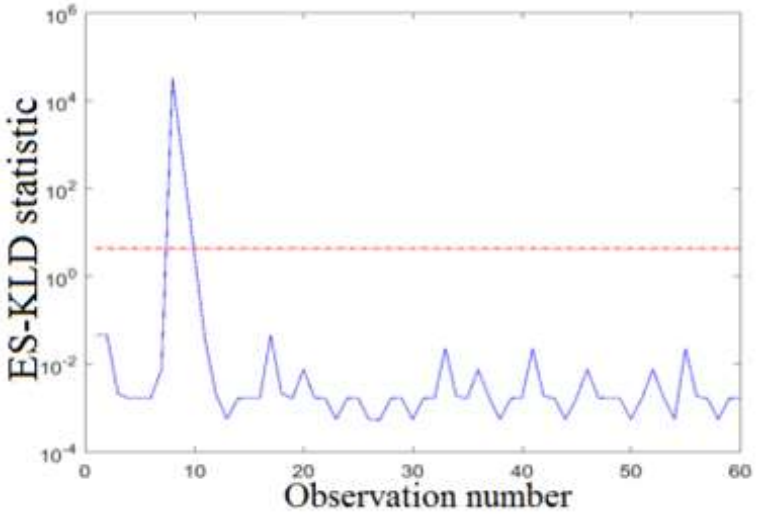

Figure 5.22 : Résultat de détection en présence des attaques RA flood (Trafic ICMPv6)

Les figures 5.17, 5.18, 5.19, 5.20, 5.21 et 5.22 montrent les résultats de détection du mécanisme ES-KLD en présence d'attaques DOS basées sur ICMPv6. Comme l'illustrent ces figures, le contrôle des messages Neighbor advertisement, Neighbor Solicitation et Router advertisement a permis la détection des attaques considérées qui sont basées sur ces messages. En outre, les attaques dans les figures 5.18, 5.20 et 5.22 ont été détectées par le contrôle du trafic ICMPv6. En effet, les attaques dans cette base de données sont caractérisées par une intensité élevée avec 1681, 2516 et 2591 messages, respectivement. Les résultats du tableau 5.4 confirment la capacité de l'ES-KLD à détecter ces différents types d'attaques.

Attaque		TPR	FPR	Accuracy	AUC
NA flood	Trafic ICMP	1	0	1	1
	Flux de messages NA	1	0	1	1
NS flood	Trafic ICMP	1	0.033	0.969	0.984
	Flux de messages NS	1	0.033	0.968	0.983
RA flood	Trafic ICMP	1	0	1	1
	Flux de messages RA	1	0	1	1

Tableau 5.4: Performances de l'ES-KLD en présence des attaques DOS basées sur ICMPV6

5.4.4. Comparaison avec des travaux antérieurs

Le tableau 5.5 présente les résultats de comparaison de l'ES-KLD avec les techniques AID [113], BPN [129], et SVM [114] en présence des attaques TCP SYN flood de la base DARPA 99. Des TPR et FPR obtenus, on constate que les meilleures performances sont celles de l'ES-KLD. Il a permis une bonne précision de détection par rapport à ces trois techniques.

Approche	TPR (%)	FPR (%)
ES-KLD	100	0.18
AID 40x40	96.80	2.85
AID 30x30	96.30	3.15
BPN	96.30	0.70
SVM	99.20	0.84

Tableau 5.5: Comparaison avec d'autres approches de détection (scénario d'attaque DARPA 99 SYN flood)

5.5. Conclusion

Dans ce chapitre nous avons introduit l'approche ES-KLD pour la détection des attaques DOS et DDOS à base d'anomalies. Par rapport aux travaux antérieurs qui ont exploité la divergence KLD pour résoudre différents problèmes de sécurité, nous avons

proposés de nouvelles fonctionnalités pour améliorer ses performances et la rendre plus robuste et plus adaptée à la détection temps réel. En fait, ES-KLD utilise le lissage exponentiel ES des mesures KLD pour quantifier les écarts des trafics anormaux par rapport au trafic de référence. Ensuite, des attaques sont révélées si les mesures ES-KLD dépassent un seuil de décision non-paramétrique établit via l'estimation de la densité du noyau KDE.

Pour valider son efficacité, nous avons testé ES-KLD en présence des attaques TCP SYN flood, UDP flood, SMURF et les attaques à base ICMPv6, tout en considérant trois bases de données de trafic IP.

Les résultats de détection obtenus illustrent les performances élevées de l'ES-KLD. Par rapport à d'autres solutions, il présente un taux de détection élevé et réduit considérablement les fausses alarmes.

Conclusion générale et Perspectives

Conclusion générale et Perspectives

Le travail présenté dans cette thèse porte sur l'utilisation des cartes de contrôle statistiques et les mesures de similarité pour la détection des anomalies de trafic réseau, et en particulier celles liées aux cyber-attaques de dénie de service DOS et DDOS.

- **Contributions et résultats**

Pour mettre en évidence leur utilité dans la détection des attaques DOS et DDOS, nous avons introduit une étude comparative entre les cartes de contrôle Shewhart, CUSUM et EWMA. Les performances des trois cartes sont évalués sous différents scénarios d'attaques TCP SYN flood et SMURF fournis par la base de trafic IP DARPA99. Les résultats obtenus montrent que ces cartes peuvent avoir un grand intérêt dans ce sens. Notamment, la carte EWMA qui a présenté les meilleures performances.

Ensuite, nous avons développé de techniques de détection d'anomalies à base de la distance CRPS pour la détection des différents types d'attaques DOS et DDOS. Dans ces techniques, la distance CRPS est utilisée pour quantifier la déviation du trafic surveillé par rapport au trafic normal de référence. En premier lieu, nous avons proposé les cartes CRPS-Shewhart et CRPS-EWMA pour la détection paramétrique des anomalies dans des processus Gaussiens. Dans ceux approches, les mesures CRPS obtenues sert de données d'entrée aux cartes Shewhart et EWMA. Ainsi, nous avons mis en place le mécanisme CRPS-ES pour une détection automatique avec un seuil non-paramétrique, flexible et indépendant de la condition de normalité. CRPS-ES est bien adaptée à la détection temps réel, et le lissage exponentiel ES est appliqué aux mesures CRPS pour augmenter le taux de détection. Les résultats de détection montrent que l'intégration du CRPS avec Shewhart et EWMA a permis de renforcer leur sensibilité et d'améliorer le taux de détection. En fait, CRPS-EWMA a présenté les meilleures performances. Quant à CRPS-ES, la validation c'était en présence de plusieurs types des attaques DOS et DDOS dans les réseaux IPv4 et IPv6. Les résultats obtenus montrent l'efficacité du mécanisme CRPS-ES à révéler les attaques considérées, en offrant de

taux de détection élevés et de faibles fausses alarmes. Comparé à d'autres travaux, les meilleures performances sont obtenues avec CRPS-ES.

Pour remédier aux limitations des approches de détection basées sur la divergence KLD, nous avons proposé une nouvelle implémentation appelée ES-KLD. Les nouveautés dans ES-KLD concernent son adaptation aux applications de détection temps réel, détection automatique des anomalies, fonctionnement indépendant de la distribution du trafic via KDE et amélioration de la sensibilité par le lissage ES. En évaluant ses performances sous différents scénarios de cyber-attaques DOS et DDOS, ES-KLD a permis de résultats de détection prometteuses.

- **Perspectives :**

Avec les résultats de détection prometteurs des CRPS-ES et ES-KLD, nous avons l'intention d'étendre le présent travail pour considérer d'autres fromes d'attaques DDOS, à savoir les attaques LR-DOS/DDOS dans un trafic réel hors ligne ou sur un réseau opérationnel, en ligne. D'autre part, afin d'améliorer encore les performances de CRPS-ES et l'ES-KLD en termes de précision de détection et de réduction du taux de fausses alarmes, nous prévoyons d'inclure d'autres caractéristiques de trafic (comme les segments TCP ACK/RST/FIN et les ports) dans le processus de détection.

Récemment, les techniques de machine et deep learning constituent un axe de recherche prometteur à bord de nombreuses disciplines. Dans des travaux futurs, nous prévoyons de combiner des réseaux neuronaux récurrents (RNN) dérivés de modèles de deep learning qui tiennent compte des dépendances temporelles intrinsèquement et non linéaire des caractéristiques du trafic avec la sensibilité des CRPS et KLD pour la détection efficace des intrusions.

Références bibliographiques

Références bibliographiques

[1] S. Bhatia, Detecting Distributed Denial-of-Service Attacks and Flash Events, Phd thesis, Queensland University of Technology, 2013.

[2] M. Thottan and C. Ji, Anomaly Detection in IP Networks, IEEE Transactions on Signal Processing, Vol. 51, no. 8, 2003.

[3] M. Graham, W. H. Dutton, Society and the Internet: How Networks of Information and Communication Are Changing Our Lives, Oxford University Press, 2019.

[4] Arbor networks : https://www.crunchbase.com/organization/arbor-networks/timeline/timeline#section-recent-news-activity

[5] O. Kupreev, E. Badovskaya and A.Gutnikov, DDoS attacks in Q2 2020, Kaspersky, 2020.

[6] M.V.K. Ronning, Mitigating DDoS attacks using data mining and density-based geographical clustering, Master thesis, University of Oslo, 2017.

[7] M. Handley, E. Rescorla, Internet Denial-of-Service Considerations, RFC 4732 November 2006.

[8] R. Kesavamoorthy, P. Alaguvathana, R. Suganya and P. Vigneshwaran, Classification of DDoS Attacks : A Survey, Test Engineering and Management journal, Vol. 83, 2020.

[9] W. Zhijun, L. Wenjing, L. Liang and Y. Meng, Low-Rate DoS Attacks, Detection, Defense, and Challenges: A Survey, IEEE Access, Vol. 8, pp. 43920-43943, 2020.

[10] T. Mahjabin, Y. Xiao, G. Sun and W. Jiang, A survey of distributed denial-of-service attack, prevention, and mitigation techniques, International Journal of Distributed Sensor Networks, Vol. 13(12), 2017.

[11] S. Deore and A. Patil, Survey Denial of Service classification and attack with Protect Mechanism for TCP SYN Flooding Attacks, International Research Journal of Engineering and Technology, Vol. 3(5), pp. 1736-1739, 2016.

[12] M. Masdari and M. Jalali, A survey and taxonomy of DoS attacks in cloud Computing, Security and Communication Networks Journal, Vol. 9(16), pp. 3724-3751, 2016.

[13] H. Wang, D. Zhang, and K. G. Shin, Detecting SYN Flooding attacks, in the proceedings of Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, pp. 1530–1539, 2002.

[14] H. Salunkhe, S. Jadhav, and V. Bhosale, Analysis and review of TCP SYN flood attack on network with its detection and performance metrics, International Journal of Engineering Research and Technology, Vol. 6(1), pp. 250-256, 2017.

[15] M. Bogdanoski, T. Shuminoski, and A. Risteski, Analysis of the SYN flood DoS attack, International Journal of Computer Network and Information Security, Vol. 5, no. 8, pp. 1-11, 2013.

[16] J. David and C. Thomas, Efficient DDoS flood attack detection using dynamic thresholding on flow-based network traffic, Computers and Security, Vol. 82, pp 284-295, 2019.

[17] M. Azahari, M.Yusof, F. Hani, M. Ali, and M.Y. Darus, Detection and Defense Algorithms of Different Types of DDoS Attacks, International Journal of Engineering and Technology, Vol. 9, no. 5, pp. 410-414, 2017.

[18] O. E. Elejla, M. Anbar, and B. Belaton, ICMPv6-based DoS and DDoS attacks and defense mechanisms, IETE Technical Review, Vol. 34, no. 4, pp. 390-407, 2017.

[19] B. Bouyeddou , F. Harrou, B. Kadri and Y. Sun, Detecting network cyber-attacks using an integrated statistical approach, Cluster Computing The Journal of Networks, Software Tools and Applications (2020), DOI: https://doi.org/10.1007/s10586-020-03203-1.

[20] A. Conta and M. Gupta, Internet control message protocol (icmpv6) for the internet protocol version 6 (ipv6) specification, 2006.

[21] A. S. A. M. Sid Ahmed, R. Hassan and N. E. Othman, IPv6 Neighbor Discovery Protocol Specifications, Threats and Countermeasures: A Survey, IEEE access, Vol. 5, pp. 18187-18210, 2017.

[22] A. A. Bahashwan, M. Anbar and S.M. Hanshi, Overview of IPv6 Based DDoS and DoS Attacks Detection Mechanisms. In: Internation Conference on Advances in Cyber Security (ACeS 2019), pp. 153-167, 2019.

[23] P. G. Teodoroa, J. Diaz-Verdejo, G. Macia-Fernndeza, E. Vazquez, Anomaly-based network intrusion detection: Techniques, systems and challenges, Computers and Security, Vol. 28, pp. 18-28, 2009.

[24] O.E. Elejla, M. Anbar, B. Belaton and B.O. Alijla, Flow-Based IDS for ICMPv6-Based DDoS Attacks Detection, Arabian Journal for Science and Engineering, Vol. 43, pp. 7757-7775, 2018.

[25] F. Harrou, Y. Sun, A. S. Hering and M. Madakyaru, Statistical process monitoring using advanced data-driven and deep learning approaches: theory and practical applications. Elsevier, 2020.

[26] V.Chandola, A. Bnerjee and V. Kumar, Anomaly Detection : A Survey, ACM Computing Surveys, 2009.

[27] M. Ahmed, A. N. Mahmood and J.Hu, A survey of network anomaly detection techniques, Journal of Network and Computer Applications, Vol. 60, pp. 9-31, 2016.

[28] S. N. Shirazi et al., Evaluation of Anomaly Detection techniques for SCADA communication resilience, Resilience Week (RWS), pp. 140-145, 2016.

[29] R. Chalapathy and S. Chawla, deep learning for anomaly detection: a survey, 2019.

[30] S. Pukkawanna, Unsupervised anomaly Detection in Massive Traffic Using S-Transform and Rény Divergence, Phd thesis, Nara Institute of Science and Technologie, 2015.

[31] M. H. Bhuyan, D. K. Bhattacharyya and J. K. Kalita, Network Anomaly Detection: Methods, Systems and Tools, IEEE Communications Surveys and Tutorials, Vol. 16, no. 1, pp. 303-336, 2014.

[32] M. V. Joshi, R. C. Agarwal, and V. Kumar, Mining needle in a haystack: classifying rare classes via two-phase rule induction, in the Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 293-298, 2001.

[33] J. Theiler and D. M. Cai, Resampling approach for anomaly detection in multispectral images, in the Proceedings of SPIE, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX, Vol. 5093. pp. 230-240, 2003.

[34] L. Portnoy, E. Eskin, and S. J. Stolfo, Intrusion detection with unlabeled data using clustering, In the Proceedings of The ACM Workshop on Data Mining Applied to Security, 2001.

[35] F. Harrou, L. Fillatre, M. Bobbia and I. Nikiforov, Statistical detection of abnormal ozone measurements based on constrained generalized likelihood ratio test. in the Proceedings of. the 52nd IEEE Conference on Decision and Control, pp. 4997-5002, 2013.

[36] H. H. Nguyen, N. Harbi, and J. Darmont, An efficient local region and clustering-based ensemble system for intrusion detection, In the Proceedings. of the 15th Symposium on International Database Engineering and Applications, pp. 185-191, 2011.

[37] S. Axelsson, The base-rate fallacy and the difficulty of intrusion detection, ACM Transactions on Information and System Security, Vol. 3, no. 3, pp. 186-205, 2000.

[38] Y. Wang, Statistical Techniques for Network Security : Modern Statistically-Based Intrusion Detection and Protection. Hershey, PA: Information Science Reference, IGI Publishing, 2008.

[39] N. Moustafa, J. Hu and J. Slay, A holistic review of Network Anomaly Detection Systems: A comprehensive Survey, Journal of Network and Computer Applications, Vol. 128, pp.33-55, 2019.

[40] A. R. Vasudevan, E. Harshini and S. Selvakumar, SSENet-2011: A Network Intrusion Detection System dataset and its comparison with KDD CUP 99 dataset, In the Proceedings of the Second Asian Himalayas International Conference on Internet (AH-ICI), pp. 1-5, 2011.

[41] P. A. A. Resende, A. C. Drummond, A survey of random forest based methods for intrusion detection systems. ACM Computing. Surveys, Vol 51(3), 48, 2018.

[42] N. Moustafa, G. Creech, J. Slay, Big data analytics for intrusion detection system: statistical decision-making using finite dirichlet mixture models. In: Data Analytics and Decision Support for Cybersecurity, Springer, pp. 127‑156, 2017.

[43] E. Eskin, Anomaly detection over noisy data using learned probability distributions, In the Proceeding of the 7th International Conference on Machine Learning, pp. 255-262, 2000.

[44] M. Desforges, P. Jacob, and J. Cooper, Applications of probability density estimation to the detection of abnormal conditions in engineering, In the Proceeding of the Institute of Mechanical Engineers, Vol. 212, pp.687-703, 1998.

[45] Z. Zhang, J. Li, C. N. Manikopoulos, J. Jorgenson, and J. Ucles, HIDE: a Hierarchical Network Intrusion Detection System Using Statistical Preprocessing and Neural Network Classification, In the Proceeding of IEEE Man Systems and Cybernetics Information Assurance Workshop, 2001.

[46] Z. Chen, C. K. Yeo, B. S. L. Francis and C. T. Lau, Combining MIC feature selection and feature-based MSPCA for network traffic anomaly detection, Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC), pp. 176-181, 2016.

[47] H. Altwaijry, Bayesian based intrusion detection system. IAENG Transactions on Engineering Technologies, Springer, Vol. 170, pp. 29-44, 2016.

[48] S. Dua and X. Du, Data Mining and Machine Learning in Cybersecurity, first ed, Vol. 1. CRC press. 2011.

[49] C. Gruhl, B. Sick, A. Wacker, S. Tomforde, and J.Hahner, A building block for awareness in technical systems: online novelty detection and reaction with an application in intrusion detection. In 7th International Conference on Awareness Science and Technology (iCAST), pp. 194-200, 2015.

[50] H. Shahriar and M. Zulkernine, Information-Theoretic Detection of SQL Injection Attacks, In: 14th International Symposium on High-Assurance Systems Engineering, pp. 40-47, 2012.

[51] H. Shahriar, S. North, W-C. Chen, and E. Mawangi, Information Theoretic XSS Attack Detection in Web Applications, International Journal of Secure Software Engineering, Vol. 5(3), pp. 1-15, 2014.

[52] V. N. Cooper, H. M. Haddad and H. Shahriar, Android Malware Detection Using Kullback-Leibler Divergence, Advances in Distributed Computing And Artificial Intelligence Journal, Vol. 3, no. 2, pp. 17-25, 2014.

[53] K. Ozonat, An information-theoretic approach to detecting performance anomalies and changes for large-scale distributed web services, In International Conference on Dependable Systems and Networks With FTCS and DCC (DSN), pp. 522-531, 2008.

[54] J. Tang, Y. Cheng and C. Zhou, Sketch-Based SIP Flooding Detection Using Hellinger Distance, In: IEEE Global Telecommunications Conference, pp. 1-6, 2009.

[55] W. Lu and H. Tong, Detecting Network Anomalies Using CUSUM and EM Clustering, In the Proceddings of 4th International Symposium on Advances in Computation and Intelligence, pp. 297-308, 2009.

[56] M. A. Qadeer, A. Iqbal, M. Zahid, and M. R. Siddiqui, Network Traffic Analysis and Intrusion Detection Using Packet Sniffer, In the Proceedings of 2nd International Conference on Communication Software and Networks, pp. 313-317, 2010.

[57] I. Kang, M. K. Jeong, and D. Kong, A differentiated one-class classification method with applications to intrusion detection, Expert Systems with Applications, Vol. 39, no. 4, pp. 3899-3905, 2012.

[58] C. F. Tsai, Y. F. Hsu, C. Y. Lin, and W. Y. Lin, Intrusion detection by machine learning: A review, Expert Systems with Applications, Vol. 36, no. 10, pp. 11994-12000, 2009.

[59] C. Wagner, J. Franc¸ois, R. State, and T. Engel, Machine Learning Approach for IP-Flow Record Anomaly Detection, In the Proceedings of 10th International IFIP conference on Networking , pp. 28-39, 2011.

[60] B. Scholkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson, Estimating the Support of a High-Dimensional Distribution, Neural Computation, Vol. 13, no. 7, pp. 1443-1471, 2001.

[61] S. J. Horng, M. Y. Su, Y. H. Chen, T. W. Kao, R. J. Chen, J. L. Lai, and C. D. Perkasa, A novel intrusion detection system based on hierarchical clustering and support vector machines. Expert Systems Application, Vol. 38 (1), pp. 306-313, 2011.

[62] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, Network traffic anomaly detection techniques and systems, In: Network Traffic Anomaly Detection and Prevention. Springer, pp. 115‑169, 2017.

[63] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, Nado: network anomaly detection using outlier approach. In the Proceedings of the International Conference on Communication, Computing and Security, ACM, pp. 531-536, 2011.

[64] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, An effective unsupervised network anomaly detection method, In the Proceedings of the International Conference on Advances in Computing, Communications and Informatics, ACM, pp. 533-539, 2012.

[65] Z. Zhuang, Y. Li, and Z. Chen, Enhancing Intrusion Detection System with proximity information, International Journal of. Security and Networks, Vol. 5, no. 4, pp. 207-219, 2010.

[66] M. E. Otey, A. Ghoting, and S. Parthasarathy, Fast distributed outlier detection in mixed-attribute data sets, Data Mining and Knowledge Discovery, Vol. 12, no. 2-3, pp. 203-228, 2006.

[67] M. S. A. Khan, Rule based Network Intrusion Detection using Genetic Algorithm, International Journal of Computer Applications, Vol. 18, no. 8, pp. 26-29, 2011.

[68] A. Tajbakhsh, M. Rahmati, and A. Mirzaei, Intrusion detection using fuzzy association rules, Applied Soft Computing, Vol. 9, no. 2, pp. 462-469, 2009.

[69] A. Visconti and H. Tahayori, Artificial immune system based on interval type-2 fuzzy set paradigm, Applied Soft Computing, Vol. 11, no. 6, pp. 4055-4063, 2011.

[70] S. Noel, D. Wijesekera, and C. Youman, Modern Intrusion Detection, Data Mining, and Degrees of Attack Guilt, In the Proceedings of International Conference on Applications of Data Mining in Computer Security, 2002.

[71] R. Sekar, A. Gupta, J. Frullo, T. Shanbhag, A. Tiwari, H. Yang, and et al., Specification-based anomaly detection: a new approach for detecting network intrusions, In the Proceedings of ACM conference on Computer and Communications Security, pp. 265-274, 2002.

[72] X. Xu, Sequential anomaly detection based on temporal-difference learning: Principles, models and case studies, Applied Soft Computing, Vol. 10, no. 3, pp. 859-867, 2010.

[73] A. Prayote, Knowledge Based Anomaly Detection, Phd dissertation, School of Computer Science and Egineering, the University of New South Wales, 2007.

[74] K. Chadha, S. Jain, Hybrid genetic fuzzy rule based inference engine to detect intrusion in networks, In: Intelligent Distributed Computing, Springer, Vol. 321, pp. 185-198, 2015.

[75] H. Holm, Signature based intrusion detection for zero-day attacks:(not) a closed chapter? In: 47th Hawaii International Conference on System Sciences (HICSS). IEEE, pp. 4895-4904, 2014.

[76] I. Corona, G. Giacinto, and F. Roli, Adversarial attacks against intrusion detection systems: taxonomy, solutions and open issues, Information Sciences, Vol. 239, pp. 201-225, 2013.

[77] B. Jasiul, M. Szpyrka, J. Sliwa, Malware behavior modeling with colored petri nets, In: IFIP International Conference on Computer Information Systems and Industrial Management, Springer, pp. 667-679, 2014.

[78] D. Midi, A. Rullo, A. Mudgerikar and E. Bertino, KalisU a system for knowledge-driven adaptable intrusion detection for the internet of things, In: 37th International Conference on Distributed Computing Systems (ICDCS), IEEE, pp. 656-666, 2017.

[79] N. G. Duffield, P. Haffner, B. Krishnamurthy, and H. Ringberg, Rule-Based Anomaly Detection on IP Flows, In the Proceedings of 28th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, pp. 424-432, 2009.

[80] R. E. Schapire, A brief introduction to boosting, In the Proceedings of 16th International Joint Conference on Artificial Intelligence, pp. 1401-1406, 1999.

[81] S. S. Hung and D. S. M. Liu, A user-oriented ontology-based approach for network intrusion detection, Computer Standards and Interfaces, Vol. 30, no. 1-2, pp. 78-88, 2008.

[82] F. Harrou, B. Taghezouit and Y. Sun, Improved k NN-Based Monitoring Schemes for Detecting Faults in PV Systems. IEEE Journal of Photovoltaics, Vol. 9(3), pp. 811-821, 2019.

[83] D.C. Montgomery, Introduction to Statistical Quality Control, 6th ed., Wiley, 2009.

[84] W. H. Woodall, B. M. Adams, J. C. Benneyan, The Use of Control Charts in Healthcare, eds., Wiley, 2011.

[85] G. Suman and D.R. Prajapati, Control chart applications in healthcare: a literature review, International Journal of Metrology and Quality Engineering, Vol. 9, 5, 2018.

[86] M. Kovarik, L. Sarga and P. Klímek, Usage of control charts for time series analysis in financial management, Journal of Business Economics and Management, Vol. 16, pp. 138-158, 2015.

[87] T. H. D. Nguyen, Using Control Charts for Detecting and Understanding Performance Regressions in Large Software, In: 5th International Conference on Software Testing, Verification and Validation, pp. 491-494, 2012.

[88] H. Saulo, V. Leiva and F. Ruggeri, Monitoring Environmental Risk by a Methodology Based on Control Charts, In: Kitsos C., Oliveira T., Rigas A., Gulati S. (eds) Theory and Practice of Risk Assessment. Springer, Vol. 136. pp. 177-197, 2015.

[89] F. Harrou, B. Khaldi, Y. Sun and F. Cherif, An efficient statistical strategy to monitor a robot swarm, IEEE Sensors Journal, Vol. 20(4), pp.2214-2223, 2019.

[90] H. J. Lenz, G. B. Wetherill and P. T. Wilrich, Frontiers in Statistical Quality Control, Springer, 2013.

[91] M. Graham, Theory and applications of univariate distribution-free Shewhart, CUSUM and EWMA control charts, Phd thesis, University of Pretoria, 2008.

[92] A. Zeroual, F. Harrou, Y. Sun, and N. Messai, Monitoring road traffic congestion using a macroscopic traffic model and a statistical monitoring scheme, Sustainable Cities and Society, Vol. 35, pp. 494-510, 2017.

[93] E. S. Page, Cumulative Sum Charts, *Technometrics* 3, pp 1-9, 1961.

[94] S. W. Roberts, Control chart tests based on geometric moving averages, Technometrics 1, pp. 239-250, 1959.

[95] https://www.tcpdump.org/

[96] https://www.wireshark.org/

[97] https://www.solarwinds.com/network-performance-monitor/use-cases/deep-packet-inspection

[98] https://www.ll.mit.edu/ideval/data/1999data.html

[99] E. P. Grimit, T. Gneiting, V. Berrocal, and N. A. Johnson, The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, Vol. 132, no. 621C, pp. 2925-2942, 2006.

[100] J. E. Matheson and R. L. Winkler, Scoring rules for continuous probability distributions, Management science, Vol. 22, no. 10, pp. 1087-1096, 1976.

[101] E. Kung, S. Dey, and L. Shi, The performance and limitations of n-stealthy attacks on higher order systems, IEEE Transactions on Automatic Control, Vol. 62, no. 2, pp. 941-947, 2016.

[102] C.-Z. Bai, F. Pasqualetti, and V. Gupta, Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs, Automatica, Vol. 82, pp. 251-260, 2017.

[103] B. Bouyeddou, F. Harrou, Y. Sun and B. Kadri, An Effective Network Intrusion Detection Using Hellinger Distance-Based Monitoring Mechanism, International Conference on Applied Smart Systems (ICASS), pp. 1-6, 2018.

[104] J. Tajer, A. Makke, O. Salem and A. Mehaoua, A Comparison Between Divergence Measures for Network Anomaly Detection, 7th International Conference on Network and Service Management (CNSM 11), pp.1-5, 2011.

[105] E. S. Gardner Jr, Exponential smoothing: The state of the art, Vol. 4(1), pp.1-28, 1985.

[106] B. Bouyeddou, F. Harrou, Y. Sun and B. Kadri, CRPS-based Monitoring Charts: Application to Cyber-attacks Detection, 2nd International Conference on automatic Control, Telecommunications and Signals (ICATS'17), 2017.

[107] B. Bouyeddou B. Kadri, F. Harrou and Y. Sun, DDOS-attacks detection using an efficient measurement-based statistical mechanism, Engineering Science and Technology, an International Journal, Vol. 23(4), pp 870-878, 2020.

[108] E. Martin and A. Morris, Non-parametric confidence bounds for process performance monitoring charts, Journal of Process Control, Vol. 6, no. 6, pp. 349-358, 1996.

[109] A. R. Mugdadi and I. A. Ahmad, A bandwidth selection for kernel density estimation of functions of random variables, Computational Statistics and Data Analysis, Vol. 47, no. 1, pp. 49-62, 2004.

[110] http://www.fukuda-lab.org/mawilab/data.html.

[111] O. E. Elejla, B. Belaton, M. Anbar, and A. Alnajjar, A reference dataset for icmpv6 flooding attacks, Journal of Engineering and applied sciences, Vol. 11, no. 3, pp. 476-481, 2016.

[112] https://www.gns3.com

[113] J. Zheng, H. Hu, An anomaly intrusion detection system based on vector quantization, IEICE transactions on Information Systems, Vol. 89(1), pp. 201-210, 2006.

[114] C. D. McDermott, A. Petrovski, Investigation of computational intelligence techniques for intrusion detection in wireless sensor networks, International journal of computer networks and communications, Vol. 9(4), pp. 45-56, 2017.

[115] N. Singh, and R. Agrawal, Combination of kullback–leibler divergence and manhattan distance measures to detect salient objects, Signal, Image and Video Processing, Vol. 9 (2), pp. 427-435, 2015.

[116] A. Karine, A. Toumi, A. Khenchaf, and M. El Hassouni, Target recognition in radar images using weighted statistical dictionary-based sparse representation, EEE Geoscience and Remote Sensing Letters, Vol.14 (12), pp. 2403-2407, 2017.

[117] A. Zeroual, F. Harrou, Y. Sun, and N, Messai, Integrating model-based observer and kullback–leibler metric for estimating and detecting road traffic congestion, IEEE Sensors. Journal. Vol. 18(20), pp. 8605–8616, 2018.

[118] F. Harrou, Y. Sun and M. Madakyaru, Kullback-leibler distance-based enhanced detection of incipient anomalies. Journal of Loss Prevention in the Process Industries, Vol. 44, pp. 73–87, 2016.

[119] A. S. Leonard, D. Weissman, B. Greenbaum, E. Ghedin and K. Koelle, Transmission bottleneck size estimation from pathogen deep-sequencing data, with an application to human influenza A virus. Journal of Virology, Vol. 91(14), 2017.

[120] H. Li, J. Zhang, and X. He, Design of data-injection attacks for cyberphysical systems based on kullback-leibler divergence, Neurocomputing Vol. 361, pp. 77-84, 2019.

[121] Q. Zhang, K. Liu, Y. Xia, and A. Ma, Optimal stealthy deception attack against cyber-physical systems, IEEE Transactions on Cybernetics, Vol. 50, no. 9, pp. 3963-3972, 2019.

[122] J. E. Tapiador, J. A. Clark, Masquerade mimicry attack detection: A randomised approach, Computers and Security, Vol. 30 (5), pp. 297-310, 2011.

[123] B. Bigi, Using Kullback-Leibler Distance for Text Categorization, In: Sebastiani F. (eds) Advances in Information Retrieval. ECIR Lecture Notes in Computer Science, Vol. 2633. Springer, 2003.

[124] G. Li and Y. Wang, Differential Kullback-Leibler Divergence Based Anomaly Detection Scheme in Sensor Networks, 2012 IEEE 12th International Conference on Computer and Information Technology, pp. 966-970, 2012.

[125] D. Brauckhoff and X. Dimitropoulos, A. Wagner and K. Salamatian, Anomaly Extraction in Backbone Networks Using Association Rules, in IEEE/ACM Transactions on Networking, Vol. 20, no. 6, pp. 1788-1799, 2012.

[126] L. Pardo, Statistical inference based on divergence measures. Chapman and Hall/CRC, 2005.

Liste des Publications

Liste des publications

Publications internationales:

1- B. Bouyeddou B. Kadri, F. Harrou and Y. Sun, DDOS-attacks detection using an efficient measurement-based statistical mechanism, Engineering Science and Technology, an International Journal, Vol. 23 (issue 4), pp. 870–878, August 2020. DOI: https://doi.org/10.1016/j.jestch.2020.05.002.

2- B. Bouyeddou , F. Harrou, B. Kadri and Y. Sun, Detecting network cyber-attacks using an integrated statistical approach, Cluster Computing the Journal of Networks, Software Tools and Applications (2020), DOI: https://doi.org/10.1007/s10586-020-03203-1

Communications internationales:

1- B. Bouyeddou, B. Kadri, F. Harrou and Y. Sun, Nonparametric Kullback-Leibler distance-based method for networks intrusion detection, International Conference on Data Analytics for Business and Industry (ICDABI'20), 26-27 October, 2020, Bahrain (Online).

2- B. Kadri, B. Bouyeddou and D. Moussaoui, Early Fire Detection System Using Wireless Sensor Networks, 2018 International Conference on Applied Smart Systems (ICASS'18), 24-25 November, 2018, Medea, Algeria, pp. 1-4, doi: 10.1109/ICASS.2018.8651977.

3- F. Harrou, B. Bouyeddou, Y. Sun and B. Kadri, A Method to Detect DOS and DDOS Attacks based on Generalized Likelihood Ratio Test, 2018 International Conference on Applied Smart Systems (ICASS'18), 24-25 November, 2018, Medea, Algeria, pp. 1-6, doi: 10.1109/ICASS.2018.8652030.

4- B. Bouyeddou, F. Harrou, Y. Sun and B. Kadri, An Effective Network Intrusion Detection Using Hellinger Distance-Based Monitoring Mechanism, 2018 International

Conference on Applied Smart Systems (ICASS), Medea, Algeria, 2018, pp. 1-6, doi: 10.1109/ICASS.2018.8652008.

5- B. Bouyeddou, F. Harrou, Y. Sun and B. Kadri, Detection of smurf flooding attacks using Kullback-Leibler-based scheme, 2018 4th International Conference on Computer and Technology Applications (ICCTA), 3-5 May, 2018, Istanbul, pp. 11-15, doi: 10.1109/CATA.2018.8398647.

6- B. Bouyeddou, F. Harrou, Y. Sun and B. Kadri, CRPS-based Monitoring Charts: Application to Cyber-attacks Detection, 2nd International Conference on automatic Control, Telecommunications and Signals (ICATS'17), 11-12 December 2017, Annaba, Algeria.

7- B. Bouyeddou, F. Harrou, Y. Sun and B. Kadri, Detecting SYN flood attacks via statistical monitoring charts: A comparative study, 5th International Conference on Electrical Engineering - Boumerdes (ICEE-B), 29-31 October, 2017, Boumerdes, pp. 1-5, doi: 10.1109/ICEE-B.2017.8192118.