الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة سعيدة – د. الطاهر مولاي –
Université Saïda– Dr. Tahar Moulay –
Faculté des Mathématiques, Informatique et Télécommunications

MEMOIRE

Présenté pour l'obtention du **Diplôme** de **MASTER en** Télécommunications

Spécialité : Réseaux et Télécommunications

Par :M. BELGACEM EL-habib
M. BOUANANI Mohamed

Application de la théorie des valeurs extrêmes à la détection des cyberattaques dans les réseaux IPv6

Soutenu, le 14/06/ 2026, devant le jury composé de :

M.GUENDOUZ Mohammed	MCA	Président
M.BOUYEDDOU Benamar	MCA	Rapporteur
Mme. BOUCHENAK Sofia	MCB	Examinateur

2025 / 2026

Remercîments

Au terme de ce travail, nous devons remercier tout d'abord dieu qui nous a donné la force et le courage de suivre nos études et d'arriver à ce stade et à nos parents qui nous ont beaucoup soutenus pendent tous le long de notre parcourt.

Un grand merci à mon encadreur **M.BOUYEDDOU Benamar** qui nous a beaucoup aidé, soutenu et nous a permis d'arriver à ce niveau-là et pour ces excellents conseils et surtout pour son temps passé avec nous et sa patience, sans lui on n'aurait pas pu réaliser ce modeste travail et pour sa confiance en nous.

Nous tenons aussi à remercier les membres de jury qui nous ont fait honneur d'examiner ce travail.

Enfin, nous renouvelons nos remerciements à ceux qui nous ont aidés de près ou de loin pour réaliser ce travail sans oublier les enseignants qui ont contribués à notre formation.

DEDICACE

Merci mon dieu de m'avoir donné la capacité d'écrire et de réfléchir, la patience d'aller jusqu'au bout du rêve J'ai l''honneur de dédier ce modeste travail :

A tous les membres de ma famille, petite et grande.

A mes chers amis.

A mon encadreur

M.BOUYEDDOU Benamar

Résumé

Ce mémoire étudie l'application de la théorie des valeurs extrêmes (EVT) dans la détection des cyberattaques au sein des réseaux IPv6. Cette étude est motivée par l'augmentation des attaques informatiques et les limites des méthodes classiques de détection des anomalies. Les principales caractéristiques du protocole IPv6 ainsi que le protocole ICMPv6 ont été présentés, avec un intérêt particulier pour les attaques DoS et DDoS telles que NAS Flood, NAD Flood, RAD Flood et PING Flood. Les concepts fondamentaux de la théorie des valeurs extrêmes ont ensuite été expliqués, notamment les distributions statistiques GEV et GPD utilisées pour l'analyse des valeurs rares. Dans la partie pratique, des données de trafic réseau contenant plusieurs types d'attaques ont été analysées afin de déterminer des seuils statistiques permettant la détection des comportements anormaux. Une comparaison des méthodes POT-GPD et GEV a été réalisée dans un environnement IPv6. Les résultats obtenus montrent que l'EVT améliore l'efficacité de la détection des cyberattaques tout en réduisant les fausses alertes. L'étude a également montré que la méthode POT-GPD offre de meilleures performances que la méthode GEV dans plusieurs cas de test. Enfin, ce travail confirme l'importance des approches statistiques modernes pour renforcer la sécurité des réseaux IPv6.

Abstract

This thesis studies the application of Extreme Value Theory (EVT) in the detection of cyberattacks within IPv6 networks. This study is motivated by the increasing number of cyberattacks and the limitations of traditional anomaly detection methods. The main characteristics of the IPv6 protocol and the ICMPv6 protocol were presented, with particular focus on DoS and DDoS attacks such as NAS Flood, NAD Flood, RAD Flood, and PING Flood. The fundamental concepts of Extreme Value Theory were then explained, especially the GEV and GPD statistical distributions used for analyzing rare and extreme values. In the practical part, network traffic datasets containing several types of attacks were analyzed in order to determine statistical thresholds for detecting abnormal behaviors. A comparison between the POT-GPD and GEV methods was also carried out in an IPv6 environment. The obtained results show that EVT improves the efficiency of cyberattack detection while reducing false alarms. The study also demonstrated that the POT-GPD method provides better performance than the GEV method in several test cases. Finally, this work confirms the importance of modern statistical approaches in strengthening IPv6 network security and improving intrusion detection systems.

الملخص

تهدف هذه المذكرة إلى دراسة تطبيق نظرية القيم القصوى (Extreme Value Theory - EVT) في مجال كشف الهجمات السيبرانية داخل شبكاتIPv6 ، وذلك بسبب التزايد المستمر للهجمات الإلكترونية وصعوبة اكتشاف السلوكيات غير الطبيعية في حركة المرور الشبكية باستخدام الطرق التقليدية.

في البداية، تم التطرق إلى بروتوكول IPv6 وخصائصه الأساسية مثل فضاء العنونة الكبير، وآليات التهيئة الذاتية، وتحسينات الأمن مقارنة بـ .IPv4 كما تمت دراسة بروتوكول ICMPv6 وأهم الهجمات المرتبطة به، خاصة هجمات الحرمان من الخدمة DoS و DDoSمثل NAS Flood و NAD Floodو RAD Floodو.PING Flood

بعد ذلك، تم عرض المفاهيم الأساسية لنظرية القيم القصوىEVT ، مع شرح أهم التوزيعات الإحصائية المستعملة في تحليل القيم النادرة والمتطرفة، مثل توزيعات GEV وGPD، إضافة إلى أهم طرق التقدير الإحصائي المستعملة في هذا المجال.

في الجانب التطبيقي، تم استخدام بيانات لحركة المرور الشبكية تحتوي على عدة أنواع من الهجمات الإلكترونية، ثم تطبيق نماذج EVT من أجل تحديد العتبات الإحصائية للكشف عن السلوكيات غير الطبيعية داخل الشبكة. كما تمت مقارنة أداء طريقتي POT-GPD و GEVفي كشف الهجمات المختلفة داخل بيئة.IPv6

أظهرت النتائج أن نظرية القيم القصوى توفر دقة جيدة في كشف الهجمات السيبرانية وتقليل الإنذارات الخاطئة، مع فعالية أكبر لطريقة POT-GPD في العديد من حالات الاختبار مقارنة بطريقة.GEV

وفي الأخير، تؤكد هذه الدراسة أهمية توظيف الأساليب الإحصائية الحديثة، وخاصةEVT ، لتحسين أنظمة كشف التسلل وتعزيز أمن شبكات IPv6 ضد الهجمات السيبرانية المختلفة.

Table des matières

Liste des figures

Liste des Tableaux

Liste des abréviations

- **ARP** : Address Resolution Protocol
- **BM** : Block Minima
- **DAD** : Duplicate Address Detection
- **DHCP(v6)** : Dynamic Host Configuration Protocol
- **DNS** : Domain Name System
- **DoS** : Déni de service
- **DDoS** : Déni de service distribué
- **EVT (TVE)** : Théorie des Valeurs Extrêmes
- **GEV** : Generalized Extreme Value
- **GPD** : Generalized Pareto Distribution
- **ICMP(v6)** : Internet Control Message Protocol
- **IETF** : Internet Engineering Task Force
- **IGMP** : Internet Group Multicast Protocol
- **IID** : Indépendantes Identiquement Distribuées
- **IPsec** : Internet Protocol Security
- **IPv4** : Internet Protocol version 4
- **IPv6** : Internet Protocol version 6
- **LME** : Estimation par les L-Moments
- **ME** : Estimation par les Moments
- **MLE** : Maximum de Vraisemblance
- **MTU** : Maximum Transmission Unit
- **NAD** : Neighbor Advertisement
- **NAS** : Neighbor Solicitation
- **NAT** : Network Address Translation
- **NUD** : Neighbors Unreachability Detection
- **PORT** : Peaks Over Random Threshold
- **POT** : Peaks Over Threshold
- **QoS** : Quality of Service
- **QR-EVT** : Quantile Regression for Extremes
- **RAD** : Router Advertisement
- **SEND** : SEcure Neighbor Discovery
- **SLAAC** : Stateless Address Autoconfiguration

- **SYN** : Synchronisation
- **TCP** : Transmission Control Protocol
- **UDP** : User Datagram Protocol

Introduction Générale

Introduction Générale

L'évolution rapide des technologies de l'information et de la communication a profondément transformé les infrastructures réseaux modernes. Avec l'augmentation continue du nombre d'équipements connectés à Internet et l'essor des applications numériques, les besoins en adressage IP et en performance réseau ont considérablement augmenté. Dans ce contexte, le protocole IPv6 a été introduit afin de remédier aux limitations du protocole IPv4, notamment l'épuisement des adresses IP. Grâce à son vaste espace d'adressage, ses mécanismes d'auto-configuration et ses améliorations en matière de sécurité et de qualité de service, IPv6 représente aujourd'hui une solution incontournable pour les réseaux de nouvelle génération.

Cependant, malgré les améliorations apportées par IPv6, les réseaux restent exposés à de nombreuses menaces de sécurité. Les cyberattaques, particulièrement les attaques par déni de service distribué DDoS, les inondations ICMPv6 et les attaques exploitant les mécanismes de découverte de voisinage, constituent des risques majeurs pouvant compromettre la disponibilité, l'intégrité et la confidentialité des systèmes informatiques. La détection rapide et efficace de ces attaques est devenue un enjeu essentiel pour assurer la fiabilité et la continuité des services réseaux.

Dans ce cadre, les méthodes classiques de détection d'intrusion montrent parfois leurs limites face à la complexité et au volume croissant du trafic réseau. Il devient alors nécessaire de développer des approches capables d'identifier les comportements rares et anormaux présents dans les données réseau. La théorie des valeurs extrêmes (TVE), branche de la statistique dédiée à l'étude des événements rares et extrêmes, apparaît comme une solution particulièrement adaptée à ce problème. Elle permet de modéliser les observations inhabituelles situées dans les queues de distribution et d'estimer les seuils critiques au-delà desquels une activité peut être considérée comme suspecte.

L'objectif principal de ce travail est d'étudier l'application de la théorie des valeurs extrêmes à la détection des cyberattaques dans les réseaux IPv6. Plus précisément, ce mémoire s'intéresse à l'utilisation des modèles EVT, notamment les approches GEV et POT-GPD, afin de détecter les anomalies générées par différents types d'attaques réseau. Cette étude vise également à comparer les performances de ces approches en termes de détection des événements extrêmes dans le trafic IPv6.

Pour atteindre ces objectifs, nous avons adopté une méthodologie basée sur la collecte et l'analyse de bases de données contenant différents scénarios d'attaques tels que ICMP Flood,UDP Flood, SYN Flood, SDN-DDoS, NAS, NAD et RAD. Les données ont été prétraitées puis analysées à l'aide des modèles EVT afin de déterminer des seuils de détection adaptés aux comportements anormaux du trafic réseau.

Ce mémoire est organisé en trois chapitres principaux. Le premier chapitre présente l'environnement IPv6, ses caractéristiques ainsi que les principales problématiques de sécurité associées aux réseaux IPv6. Le deuxième chapitre est consacré aux fondements théoriques de la théorie des valeurs extrêmes, aux distributions GEV et GPD ainsi qu'aux différentes méthodes d'estimation utilisées. Enfin, le troisième chapitre décrit la méthodologie adoptée, l'architecture de détection proposée, les bases de données exploitées ainsi que les résultats expérimentaux obtenus et leur interprétation.

Chapitre I :
Environnement IPv6 et les problèmes de sécurité

I.1.Introduction

L'IPv6 est un protocole IP de nouvelle génération développé pour résoudre le problème de l'épuisement des adresses et fournir une infrastructure plus évolutive, plus sûre et plus efficace pour la communication sur l'internet. Le changement le plus notable de l'IPv6 est son espace d'adressage nettement plus grand. Alors que l'IPv4 utilise des adresses de 32 bits (qui fournissent environ 4,3 milliards d'adresses uniques), l'IPv6 utilise des adresses de 128 bits, ce qui permet un nombre presque illimité d'adresses IP uniques - environ 340 sous-milliards (3,4 × 10^38. Ce chapitre permet de mettre l'accent sur le protocole IPv6 ainsi que les problèmes de sécurité dans les environnements associés.

I.2.IPv6 (Internet Protocol version 6)

I.2.1.Présentation générale

Internet Protocol version 6 ou IPv6 est la version la plus récente du protocole Internet IP. le protocole de communication qui fournit un système d'identification et de localisation pour les ordinateurs sur les réseaux et achemine le trafic sur Internet. IPv6 a été développé par l'IETF pour résoudre le problème d'épuisement des adresses IPv4. Ainsi, IPv6 est destiné à remplacer IPv4. IPv6 est devenu un projet de norme en décembre 1998 et est devenu une norme Internet le 14 juillet 2017[1].

I.2.2.Format d'entête IPV6

Le format général d'un entête IPv6 est illustré sur le Figure I.1 [2].

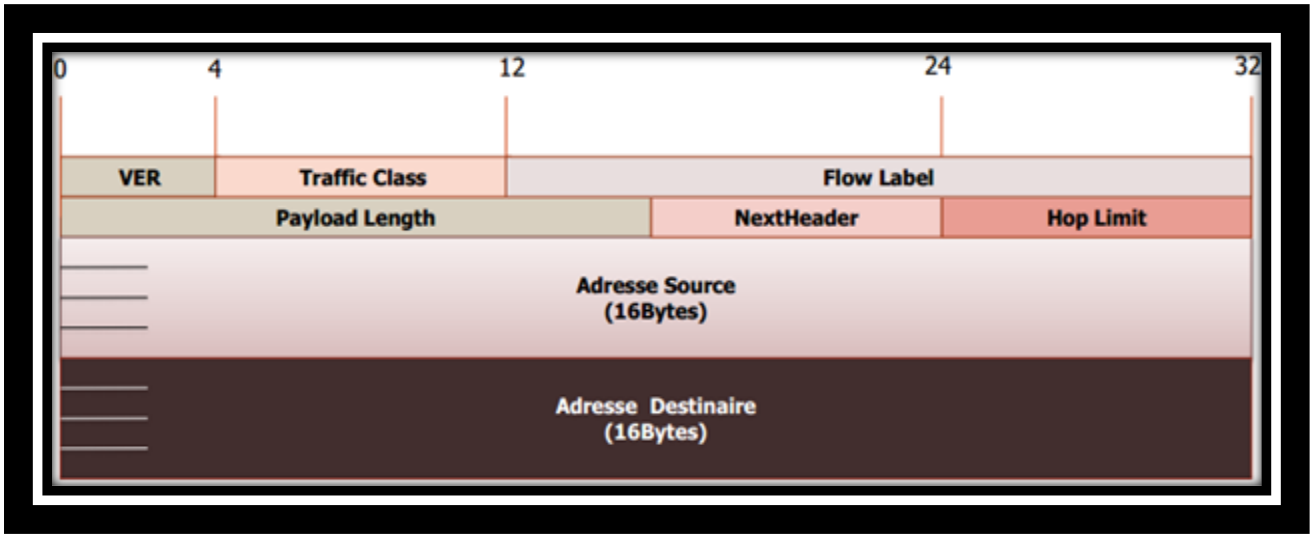

Figure I.1: Format de l'entête IPv6

- Version (4 bits) :

Numéro de version. Porte le numéro 6 (IPv6).

- Classe de trafic (8 bits):

Indique un niveau de priorité, qui permet de traiter les paquets plus ou moins rapidement dans les nœuds du réseau.

- Identificateur de flux (20 bits):

Ce champ permet de transporter une référence (label) capable de préciser le flot auquel appartient le paquet et donc d'indiquer la qualité de service exigée par les informations transportées.

- Longueur des données (16 bits) :

Indique la longueur totale du datagramme en octet, sans tenir compte de l'en-tête. Ce champ étant de 2 octets, la longueur maximale du datagramme est de 64 Ko.

- Entête suivant (8bits) :

Type de l'entête suivant, indique le protocole encapsulé dans la zone de données du paquet.

- Nombre de sauts (8 bits) :

Nombre de sauts possibles avant que le paquet ne se détruise.

- Adresse source (128 bits):

Adresse de l'élément qui envoie les données. Celle-ci est codée sur 128 bits.

- Adresse destination (128 bits):

Adresse de l 'élément qui va recevoir les données. Celle-ci est codée sur 128 bits et peut être différente si l'option routing header est activée.

- Options:

L'en-tête du paquet IPv6 se termine par un champ d'extension qui permet l'ajout de nouvelles fonctionnalités.

I.2.3.Les caractéristiques d'IPV6

Parmi les fonctionnalités les plus importantes d'IPv6, on peut citer [3] :

- Grand espace d'adresses:

L'espace d'adressage offert par IPv6 est plus grand que celui qui est fournit par IPv4. La taille de l'adresse IP passe de 32 bits pour IPv4 à 128 bits pour IPv6, soit de $4.29*10^9$ adresses IPv4 à $3.4*10^{38}$ adresses IPv6. Ceci qui permet une prise en charge d'un plus grand nombre de niveaux de hiérarchie d'adressage.

- Sécurité renforcée :

IPSec est intégré dans IPv6 dans le cadre du protocole. Cela signifie que deux périphériques peuvent créer dynamiquement un tunnel sécurisé sans intervention de l'utilisateur.

- Améliorations d'en-tête :

Avec le format d'en-tête IPv6, certains champs d'en-tête IPv4 ne sont plus utilisés tandis que d'autres deviennent facultatifs. Cette modification permet de maintenir les coûts de bande passante de l'en-tête IPv6 aussi bas que possible, malgré l'augmentation de la taille d'adresse.

- Pas besoin de NAT :

Puisque chaque périphérique a une adresse IPv6 unique au monde, il n'y a pas besoin de NAT. Auto-configuration d'adresses sans état les périphériques IPv6 peuvent se configurer automatiquement avec une adresse IPv6.

- Traitement intermédiaires allégé :

En IPv6, les routeurs intermédiaires ne fragmentent plus les paquets et renvoient un paquet ICMPv6 Packet Too Big en lieu et place, c'est alors la machine émettrice qui est responsable de fragmenter le paquet. L'utilisation du Path MTU discovery est cependant recommandée pour éviter toute fragmentation. Ce changement permet de simplifier la tâche des routeurs, leur demandant moins de puissance de traitement.

I.2.4.Phases et Méthodes de Déploiement d'IPv6

IPv6 introduit des fonctionnalités supplémentaires dans un réseau existant. Par conséquent, lors du premier déploiement d'IPv6, vous devez vous assurer de ne pas perturber les opérations fonctionnant avec IPv4 [4].

I.2.4.1.Préparation de la topologie réseau pour une prise en charge d'IPv6

La première étape du déploiement IPv6 consiste à déterminer les entités existantes sur votre réseau prenant en charge IPv6. Dans la plupart des cas, la topologie du réseau (les câbles, les routeurs et les hôtes) n'est pas modifiée par l'implémentation d'IPv6. Cependant, dans certains cas,l est nécessaire de préparer le matériel et les applications existantes pour IPv6 avant d'effectuer la configuration des adresses IPv6 sur les interfaces réseau. Assurez que le matériel de votre réseau peut être mis à niveau vers IPv6. Par exemple (routeurs, pare-feux ,serveurs ,commutateurs) [4].

I.2.4.2.Préparation des services réseau pour la prise en charge d'IPv6

Les services réseau IPv4 tell que la HTTP, DNS, ne sont pas directement compatibles avec le protocole IPv6, car les deux protocoles ont des structures d'en-tête différentes et ne peuvent pas communiquer directement, La coexistence nécessite La double pile ou Dual-

Stack est la technique la plus simple, elle consiste à doter les nœuds d'une pile IPv6 et d'une autre IPv4 ou la deuxième alternative est de créer des tunnels pour transporter IPv6 dans IPv4 ou l'inverse.

I.2.4.3.Préparation des serveurs pour une prise en charge d'IPv6

Les serveurs étant considérés comme des hôtes IPv6, par défaut, leurs adresses IPv6 sont automatiquement configurées par le protocole de détection des voisins.

I.2.4.4.Planification de tunnels dans la topologie réseau

L'implémentation d'IPv6 prend en charge un certain nombre de configurations de tunnel faisant office de mécanismes de transition lors de la migration de votre réseau vers un mélange d'IPv4 et d'IPv6. Les tunnels permettent aux réseaux IPv6 isolés de communiquer. Dans la mesure où Internet exécute essentiellement IPv4, les paquets IPv6 de votre site doivent circuler dans Internet via des tunnels ayant pour destination des réseaux IPv6, et les scénarios les plus courants d'utilisation de tunnels dans la topologie de réseau IPv6. Le FAI qui fournit des services IPv6 permet de créer un tunnel à partir du routeur de bordure du site vers le réseau du FAI.

Il est possible de gérer un réseau distribué de grande taille avec connectivité IPv4. Pour connecter les sites distribués utilisant IPv6, l est possible d'exécuter un tunnel automatique 6to4 à partir du routeur de périphérie de chaque sous-réseau.

Il est parfois impossible de mettre un routeur à niveau vers IPv6 dans l'infrastructure de l'entreprise. Dans ce cas, un tunnel manuel peut être créer à travers le routeur IPv4, avec deux routeurs IPv6 en guise d'extrémités.

I.2.5.Déploiement d'IPv6 dans le monde

Les mesures effectuées par Google montrent qu'environ 45 à 48 % des utilisateurs mondiaux se connectent désormais à ses services via IPv6. En ce début d'année 2026, la France arrive en tête avec 75,1 %, suivie de l'Inde (73,1 %) et de la Malaisie (67 %). Les Etats-Unis se situent à environ 53 à 56 %, tandis que de nombreux pays africains restent en dessous de 30 %. Selon les statistiques de Google, l'utilisation d'IPv6 a franchi la barre des 46 % au début de l'année 2026 comme illustre la figure 1.2 [5].

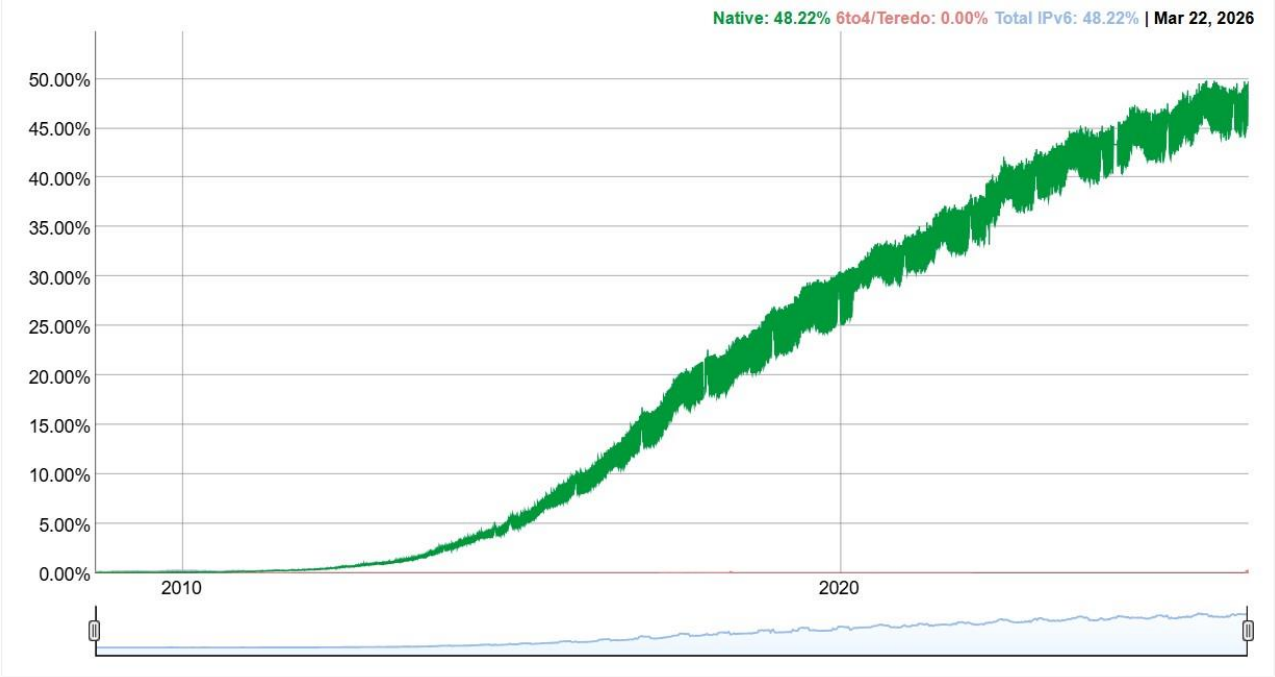

Figure I.2: Evolution du nombre d'utilisateurs google utilisant IPv6

I.3.La sécurité dans un environnement IPV6

Prise en charge native de l'authentification, de l'intégrité des données et de la confidentialité des données (par exemple, prise en charge d'IPsec). [6].

I.3.1.Améliorations apportées au protocole Internet version 6

La norme de spécification IPv6 propose certaines améliorations par rapport à IPv4. Les sous-sections suivantes fournissent de l'information supplémentaire sur les améliorations de sécurité [6].

I.3.1.1.Prise en charge du protocole IPsec

IPsec est une suite de protocoles pouvant servir à l'authentification, au chiffrement et aux mesures de protection de l'intégrité. Bien qu'il soit possible d'utiliser le protocole IPsec à titre d'extension rétroactive dans IPv4, IPv6 prend en charge IPsec, qui est intégré à la norme. Il convient de noter que le protocole IPsec n'est plus obligatoire en IPv6.

I.3.1.2.Configuration automatique

La configuration automatique permet à un nœud de s'attribuer automatiquement une adresse réseau IPv6 en fonction de l'information de préfix réseau présenté par le routeur. Pour ce faire, le mécanisme de configuration automatique d'adresse sans état SLAAC est utilisé.

I.3.1.3.Découverte de vision

Le protocole de découverte de voisin remplace le protocole de résolution d'adresse utilisé par les réseaux IPv4 afin de fournir des options de chiffrement pour des messages de découverte sécurisés.

I.3.1.4.Sécurité du protocole DHCP

Le protocole de configuration d'hôte dynamique d'IPv6 prend en charge l'authentification et le chiffrement des messages DHCP au moyen du protocole IPsec dans le but de prévenir les possibilités d'écoute clandestine et d'attaque par interception de message.

I.3.1.5.En-têtes d'extension

Les en-têtes d'extension IPv6 peuvent servir à améliorer la sécurité, au débogage et aux fonctions de gestion [6].

I.3.2.Menaces et mesures de protection

I.3.2.1.Vulnérabilité

Une vulnérabilité, appelée parfois faille, est une faiblesse dans un système informatique, laissant à un assaillant l'opportunité de porter atteinte à l'intégrité de ce système en le rendant instable. Analogiquement à une maison, cela revient à laisser une porte non-verrouillée. Du coup, cette porte (faille) peut potentiellement être utilisée par des voleurs (hackers) pour accéder de façon illégitime à la maison (au système) [7].

I.3.2.2.Menace

Une menace est une action potentielle susceptible de provoquer un dommage sur un système et ayant un impact sur ses fonctionnalités, son intégrité ou sa disponibilité. De ce fait, une vulnérabilité d'un système représente son niveau d'exposition face à une menace.

I.3.2.3.Risque

Un risque que peut provoquer une menace à un système est estimé en fonction de la sensibilité et de la vulnérabilité de ce système. Il est donc important de mesurer le risque, non seulement en se basant sur la probabilité ou sur la fréquence de son arrivée, mais aussi en mesurant son effet possible. Donc, tout système (Actif) présentant une vulnérabilité pourra être à tout moment exploitée par un assaillant. Une attaque potentielle du système est vue comme une menace visant à nuire ce système [7].

I.3.2.4.Sensibilité

Dans un système informatique, plus une information est stratégique pour une entreprise, plus elle a de la valeur, ce qui la rend sensible. De ce fait, si une information sensible est révélée au public, cela nuirait à l'entreprise et à la vie privée des individus. Concrètement, la sensibilité de l'information réside dans sa disponibilité, son intégrité et sa confidentialité.

I.3.2.5.Attaque

Une attaque est une action représentant le moyen d'exploiter une vulnérabilité pour compromettre la sécurité d'un système. Il peut y avoir plusieurs attaques pour une même vulnérabilité mais toutes les vulnérabilités ne sont pas exploitables [7].

I.3.2.6.Intrusion

Une Intrusion est un ensemble d'actions entrainant la compromission de la sécurité d'une entreprise, par exemple, accéder sans autorisation aux données du système et/ou du réseau de l'entreprise, en contournant les dispositifs de sécurité mis en place. Les raisons des intrusions sont diverses, le but pourrait être la modification ou le vol d'informations confidentielles, ou bien la destruction des données du système. Donc, l'objectif principal de la détection d'intrusion est de repérer les actions d'un attaquant qui tente de ou qui tire parti des vulnérabilités du système. Il est indispensable de définir précisément ce qui est une intrusion, c'est à dire une défaillance de sécurité.

I.3.2.7.Attaques d'un réseau

Une attaque réseau est définie comme une intrusion dans une infrastructure de communication afin d'obtenir un accès non autorisé à des ressources ou d'exploiter des vulnérabilités existantes.

I.3.3.Buts des attaques

Généralement les attaques que peut subir un réseau, cible les informations sensibles afin de nuire à l'entreprise concernée. Ils existent quatre objectifs potentiels d'une attaque :

- L'interruption : vise la disponibilité des informations dans un réseau (Déni de service, ..)
- L'interception : vise la confidentialité des données circulant dans un réseau (capture de contenu, analyse de trafic, . . .),

➢La modification : vise l'intégrité des informations en provoquant des incohérences dans le système (modification, rejeu, ...),

➢La fabrication : vise l'authenticité des informations en usurpant l'identité d'un utilisateur ou d'un service (mascarade, ...).

I.3.4.Classifications des attaques réseaux

Les attaques peuvent être classées en deux grandes catégories, selon la phase de leur opération :

➢Attaques passives : consistent à écouter et à surveiller les transmissions sur un canal, sans modifier les données dans un réseau.

➢Attaques actives: consistent à modifier les messages transmis, à s'introduire dans des équipements d'un réseau, ou à perturber le bon fonctionnement de ce dernier. Voici les attaques actives les plus connues [7].

- Par mascarade (ou usurpation d'identité)
- Par rejeu
- Par modication de message
- Par déni de service

I.3.4.1.Etapes d'une attaque

Pour procéder à une attaque d'un réseau ou un service, l'attaquant suit généralement une méthode en sept étapes :

1. Analyse d'empreinte (reconnaissance)
2. Enumération des informations
3. Manipulation des utilisateurs pour obtenir un accès
4. Escalade des privilèges
5. Collecte d'autres mots de passe et secrets
6. Installation de portes dérobées (Backdoors)
7. Exploitation du système compromise

I.3.4.2.Types d'attaques d'un réseau

Toutes les entreprises craignent les attaques informatiques, sans toujours savoir quelles formes celles-ci peuvent prendre. Au fil des années, les mesures de sécurité se sont améliorées et certains types d'attaques sont devenus moins fréquents, tandis que d'autres ont fait leur apparition. Pour concevoir des solutions de sécurité des réseaux, il faut tout d'abord

évaluer l'ensemble des attaques existantes aujourd'hui et comprendre leur mode de fonctionnement. Ces attaques sont variées, mais on peut les catégoriser selon leur mode d'opération [7]:

I.3.4.2.1.Attaques de reconnaissance

La reconnaissance est la découverte non autorisée des systèmes, de leurs adresses et de leurs services, ou encore la découverte de leurs vulnérabilités. Parmi les attaquescourantes de ce type, nous citons:

- Requêtes Internet
- Balayages ping
- Balayages de ports
- Analyse des paquets

I.3.4.2.2.Attaques d'accès

L'accès au système est la possibilité pour un assaillant d'accéder à un périphérique pour lequel il ne dispose pas d'autorisation c'est-à-dire un compte ou un mot de passe. On distingue quatre types d'attaques :

- Attaques de mot de passe
- Exploitation de la confiance
- Redirection de port
- Attaque de l'homme du milieu (man-in-the-middle)

I.3.4.2.3.Attaques par déni de service (DoS)

Le Déni de service (DoS) apparaît lorsqu'un assaillant désactive ou altère un réseau, des systèmes ou des services dans le but de refuser le service prévu aux utilisateurs normaux. Les attaques par DoS mettent le système en panne ou le ralentissent au point de le rendre indisponible et inutilisable. Il existe plusieurs types d'attaques de DoS, on peut citer [7] :

- Attaques par Ping fatal
- Attaques par Inondation SYN.
- Attaques DDoS.

I.3.4.2.4.Attaques de programmes malveillants (malware)

Des programmes/logiciels malveillants peuvent être installés sur une machine hôte dans le but d'endommager un système ou d'empêcher l'accès aux services de ce dernier, Ces

programmes essayent souvent de se reproduire et de se propager pouvant causer des dégâts multiples pour le réseau informatique, tels que : Gêne , Espionnage , Contrôle d'une machine et les virus [7].

I.4.Le protocole ICMPV6 et les attaques DOS dans un IPv6

I.4.1.Le protocole ICMPv6

I.4.1.1.Définition

Le protocole ICMPv6 est une nouvelle version du protocole ICMP qui fait partie intégrante de l'IPv6. Les messages ICMPv6 sont transportés dans un paquet IPv6 et sont identifiés par une valeur de 58 dans le champ En-tête suivant de l'en-tête IPv6 ou de l'en-tête précédent. Il assure les différentes fonctions précédemment subdivisées en différents protocoles tels qu'ICMP, ARP et IGMP [8].

ICMPv6 est un protocole polyvalent utilisé pour diverses activités, notamment la signalisation d'erreurs dans le traitement des paquets, les activités de diagnostic, le processus de découverte de voisin et la création de rapports d'appartenance à la multicast IPv6. Pour effectuer ces activités.

I.4.1.2.Types des message ICMPv6

Les messages ICMPv6 sont subdivisés en deux classes : les messages d'erreur et les messages d'information [8].

I.4.1.2.1.Les messages d'erreurs

Les messages d'erreur ICMPv6 appartiennent à quatre catégories différentes : Destination inaccessible, Temps dépassé, Paquet trop important et Problème de paramètre.

I.4.1.2.2.Les messages d'information

Les messages d'information ICMPv6 sont subdivisés en trois groupes: les messages de diagnostic, les messages de découverte de voisin et les messages de gestion des groupes de multicast [8].

I.4.1.3.Format d'un message ICMPV6

Les messages ICMPv6 ont le format indiqué sur la figureI.3, suivante [9]

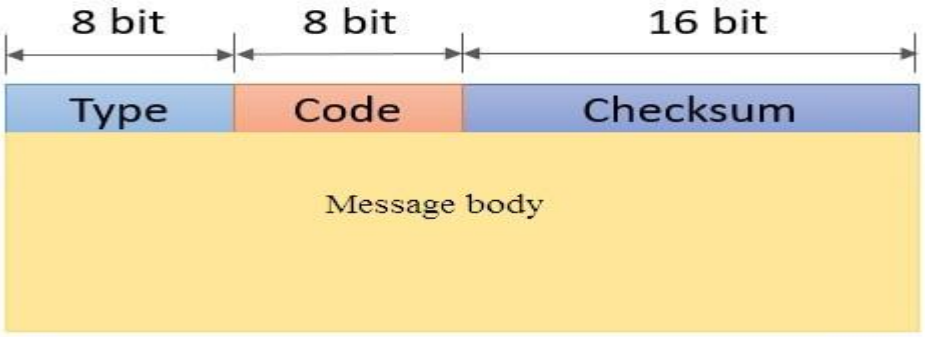

Figure I.3: format d'un message ICMPv6

- Le champ Type :

Sur 8 bits indique le type du message. Si le bit de poids fort a la valeur zéro (valeurs comprises entre 0 et 127), il indique un message d'erreur ; si le bit de poids fort a la valeur 1 (valeurs comprises entre 128 et 255), il indique un message d'information.

- Le champ Code :

De 8 bits, le contenu dépend du type de message.

- Le champ Checksum :

Champ de 16 bits, aide à détecter les erreurs dans le message ICMPv6 et dans une partie du message IPv6.

- Message Body :

Est divisé en plusieurs champs selon le type et le code du message.

I.4.1.4.Exemples de messages ICMPV6

Une liste de messages actuellement définis est indiquée dans le tableau ci-dessous [10] :

Type	Code	Nature
Gestion des erreurs		
1		Destination inaccessible
	0	Aucune route vers la destination
	1	La communication avec la destination est administrativement interdite
	2	Hors portée de l'adresse source
	3	L'adresse est inaccessible
	4	Le numéro de port est inaccessible
	5	L'adresse source est filtrée par un firewall
	6	L'adresse destination est rejetée
2		Paquet trop grand
3		Temps dépassé
	0	Limite du nombre de sauts atteinte
	1	Temps de réassemblage dépassé
4		Erreur de paramètre
	0	Champ d'en-tête erroné
	1	Champ d'en-tête suivant non reconnu
	2	Option non reconnue
Information		
128		Demande d'écho
129		Réponse d'écho
Gestion des groupes multicast (MLD, RFC 2710)		
130		Requête d'abonnement
131		Rapport d'abonnement
132		Fin d'abonnement
Découverte de voisins (RFC 2461)		
133		Sollicitation du routeur
134		Annonce du routeur
135		Sollicitation d'un voisin
136		Annonce d'un voisin
137		Redirection
Découverte de voisins sécurisée (SEND, RFC 3971)		
148		Sollicitation de chemin de certification
149		Annonce de chemin de certification

Tableau I.1 : Exemples de messages d'ICMPv6

I.4.1.5.Les attaques DOS et DDOS dans un environnement IPv6 à base ICMPv6

De nombreux attaques DOS et DDOS contre les réseaux IPv6 sont directement liés (basés) au protocole ICMPv6 en exploitant d'une part sont utilisation fréquente, et d'autre part l'absence total de mécanismes de sécurisation de ses différents messages, de la génération, l'acheminement et traitement jusqu'aux réponses des entités réceptrices. Nous présentons dans ce paragraphe quelques types d'attaques qui exploitent les messages de

Sollicitation de voisin NAS, Annonce de voisin NAD, Annonce de routeur RAD et Requête par Echo [11].

I.4.1.5.1.Attaques à base de messages NAS

I.4.1.5.1.1.Les messages NAS

Le message NAS, du tableau I.2, sert à demander des informations d'un nœud voisin, c'est-à-dire situé sur le même lien physique (ou connecté via des ponts). Le message peut lui être explicitement envoyé, ou émis sur une adresse multicast. Dans le cas de la détermination de l'adresse physique, il a la même fonction qu'une requête ARP du protocole IPv4.

Le champ adresse source du paquet IPv6 contient, soit l'adresse locale au lien, soit une adresse globale, soit l'adresse non spécifiée. Le champ adresse destination contient, soit l'adresse de multicast sollicité correspondant à l'adresse recherchée, soit l'adresse d'un nœud dans le cas d'une détection d'inaccessibilité des voisins NUD.

Le champ adresse de la cible contient l'adresse IPv6 du nœud recherché. Le champ option contient, en général, l'adresse physique de la source [11].

Type= 135	Code= 0	Total de contrôle
Réservé		
Adresse de la cible		
Option Adresse physique de la destination		

Tableau I.2: Format du message NAS

I.4.1.5.1.2.Les attaques DOS par inondation NAS

En l'absence de restriction, tous les utilisateurs, y compris les anormaux et les attaquants, peuvent poursuivre cette tâche et les destinations de ce message répondent normalement par les messages Annonce de voisin (NAD), y compris leur adresse IPv6. En conséquence, lors d'une attaque par inondation de sollicitation de voisin (figure I.4), la

victime est surchargée et se met hors service lorsqu'elle tente de répondre à toutes ces demandes.

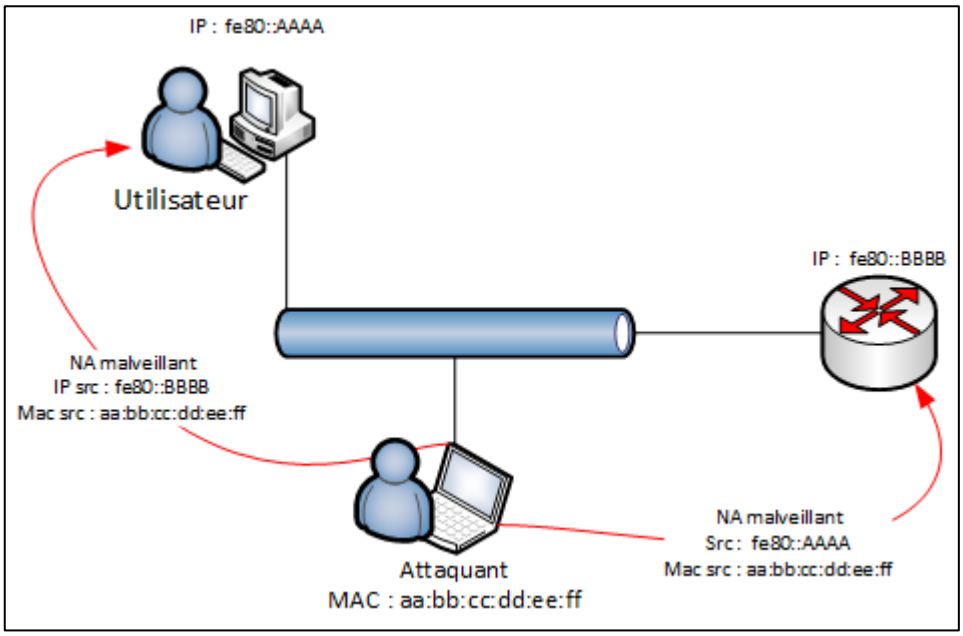

Figure I.4: Attaque DOS par messages NAS

I.4.1.5.2.Attaques à base de messages NAD

I.4.1.5.2.1.Les messages NAD

Le tableau I.3montre le format d'un message NAD. Il est émis en réponse à une sollicitation, mais il peut aussi être émis spontanément pour propager une information de changement d'adresse physique, ou de statut routeur. Dans le cas de la détermination d'adresse physique, il correspond à la réponse ARP du protocole IPv4. L'adresse de la cible, dans ce cas-là, correspond à l'adresse de la source de ce message, les champs de ce message ont la signification suivante [11] :

- Le bit R est mis à 1 si l'émetteur est un routeur.
- Le bit S mis à 1 indique que cette annonce est émise en réponse à une sollicitation.
- Le bit O mis à 1 indique que cette annonce doit effacer les informations précédentes qui se trouvent dans les caches des autres équipements, en particulier la table contenant les adresses physiques.
- Le champ adresse de la cible reprend l'adresse de la cible de la sollicitation auquel ce message répond (le bit S vaut 1 dans ce cas). Si le message d'annonce de voisin est envoyé sans sollicitation, il s'agit, pour l'émetteur, d'indiquer une nouvelle adresse

"lien-local". Le champ adresse de la cible contient alors cette nouvelle adresse "lien-local".

- L'option adresse physique de la cible contient l'adresse physique de l'émetteur.

Type= 136	Code= 0	Total de contrôle
RSO	Réservé	
Adresse de la cible		
Options (Adresse physique de la cible)		

Tableau I.3: Format du message NAD

I.4.1.5.2.2.Les attaques DOS par inondation NAD

Les messages NAS sont utilisés par le protocole DAD afin qu'une machine puisse vérifier si l'adresse qu'elle souhaite s'attribuer (grâce au mécanisme d'auto-configuration), n'est pas déjà utilisée dans le réseau. Il est possible pour un attaquant de répondre à tous les messages NAS pour annoncer qu'il possède déjà cette adresse. Cela provoque un DoS sur les machines se connectant au réseau, puisque la machine ne pourra pas s'attribuer une adresse IP et par conséquent, ne pourra pas réaliser des communications internes ou externes [11].

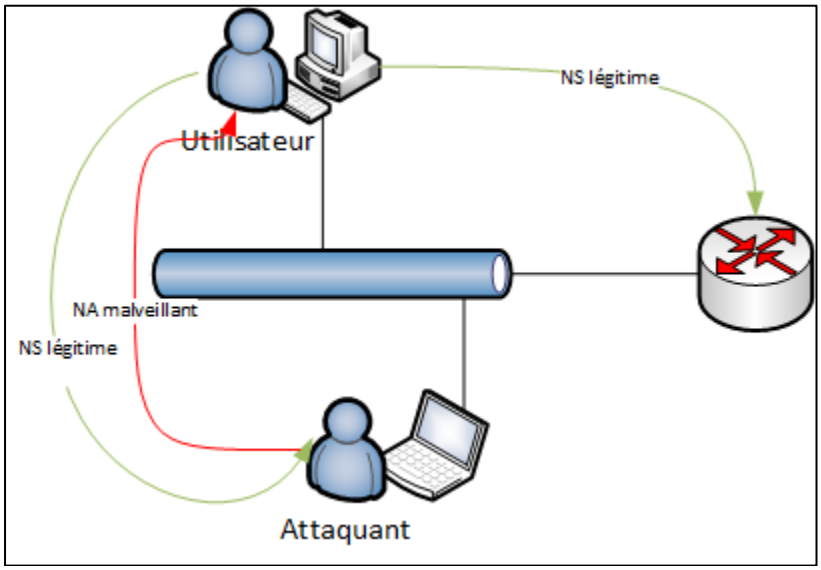

Figure I.5: Schéma d'attaque DOS en utilisant les messages NAD

I.4.1.5.3.Attaques à base de messages RAD

I.4.1.5.3.1.Les messages RAD

Ce message est émis périodiquement par les routeurs ou en réponse à un message de sollicitation d'un routeur par un équipement.

- Le champ adresse source contient l'adresse locale au lien du routeur.
- Le champ destination contient soit l'adresse de l'équipement qui a émis la sollicitation, soit l'adresse de toutes les stations (ff02::0).
- Un champ saut max. non nul donne la valeur qui pourrait être placée dans le champ nombre de sauts des paquets émis.
- Le bit M indique qu'une adresse de l'équipement doit être obtenue avec un protocole de configuration (DHCPV6).
- Le bit O indique aussi la présence d'un service de configuration mais pour la récupération d'informations autres que l'adresse. Si l'adresse ne peut être obtenue d'un serveur, l'équipement procède à une configuration sans état en concaténant aux préfixes qu'il connaît son identifiant d'interface.
- Le bit H indique que le routeur peut être utilisé comme «agentmère» pour un nœud mobile.
- Le champ duré de vie du routeur donne, en secondes, la période pendant laquelle l'équipement annonçant effectuera les fonctions de routeur par défaut. La valeur maximale correspond à 18 heures 12 minutes, mais comme ce message est émis périodiquement il n'y a pas de limite théorique à la durée de vie d'un routeur. Une valeur de 0 indique que l'équipement ne remplit pas les fonctions de routeur par défaut. Cette durée de vie ne s'applique pas aux options que ce message véhicule.
- Le champ duré d'accessibilité indique la durée en millisecondes pendant laquelle une information contenue dans le cache de la machine peut être considérée comme valide (par exemple, la table de correspondance entre adresse IPv6 et adresse physique). Au bout de cette période, un message de détection d'inaccessibilité est émis pour vérifier la pertinence de l'information.
- Le champ temporisation de retransmission donne en millisecondes la période entre deux émissions non sollicitées de ce message. Il sert aux autres équipements pour détecter une inaccessibilité du routeur.

<table>
<tr><td>Type=134</td><td colspan="4">Code=0</td><td>Total de contrôle</td></tr>
<tr><td>Cur. Limite de saut</td><td>M</td><td>O</td><td>H</td><td>Réservé</td><td>Durée de vie du routeur</td></tr>
<tr><td colspan="6">Temps accessible</td></tr>
<tr><td colspan="6">Temps de retransmission</td></tr>
<tr><td colspan="6">Options :
Adresse source de la couche liaison
MTU
Informations du prefix</td></tr>
</table>

Tableau I.4: Format des messages RAD

I.4.1.5.3.2.Les attaques DOS par inondation RAD

Sans chercher à effectuer une redirection de flux, l'attaquant peut se contenter d'envoyer des messages RAD avec le champ ROUTER LIFETIME à 0 en « spoofant » l'adresse IP du routeur légitime pour réaliser une attaque par déni de service. Dès la réception des messages « spoofés », les hôtes du réseau vont supprimer le routeur par défaut de la table de routage et ne seront plus capables d'établir des communications vers l'extérieur de leur réseau local [11].

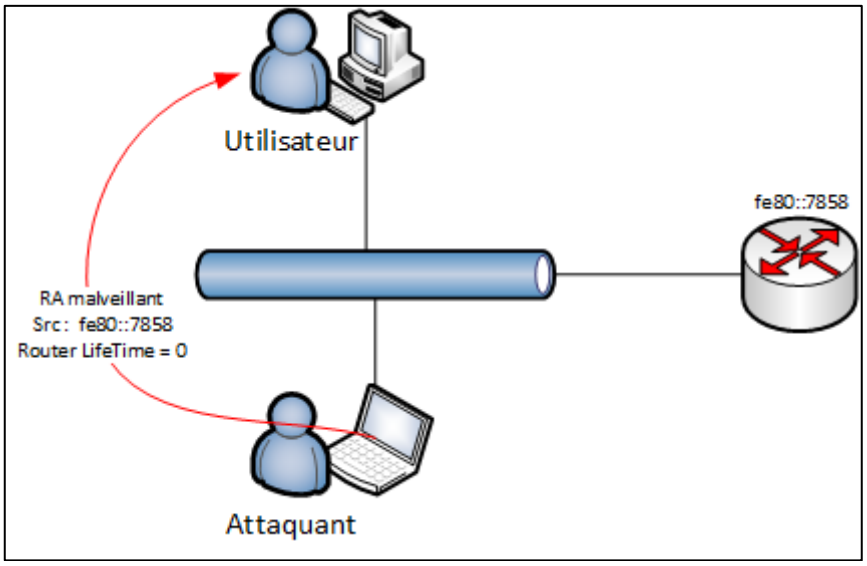

Figure I.6: Schéma d'attaque DOS en utilisant les messages RAD

I.4.1.5.4.Attaques à base de messages ECHO REQUEST (PING Flood)

I.4.1.5.4.1.Les messages ECHO REQUEST

Ces messages servent en particulier à la commande Ping permettant de tester l'accessibilité d'une machine. Le principe de fonctionnement est le même que pour IPv4, une requête (type 128) est envoyée vers l'équipement dont on veut tester le fonctionnement, celui-ci répond par le message réponse par écho (type 129). Le champ identificateur permet de distinguer les réponses dans le cas où plusieurs commandes ping seraient lancées simultanément sur la machine. Le champ numéro de séquence permet d'associer la réponse à une requête pour mesurer le temps d'aller et retour dans le cas où les demandes sont émises en continu et que le délai de propagation est élevé. Le champ donné permet d'augmenter la taille du message pour les mesures.

Type =128	Code =0	Totale de contrôle
Identificateur		Numéro de sequence
Données		

Tableau I.5: Format d'un message ICMPv6 ECHO REQUEST

I.4.1.5.4.2.Les attaques DOS par inondation ECHO REQUEST (PING Flood)

Beaucoup d'hôte Internet ou privée répondent aux paquets ICMP, il est donc facile de les inonder de ce flux afin de les rendre indisponibles. D'ailleurs, que les cibles répondent ou

pas à l'ICMP, l'objectif premier étant de saturer leurs bandes passantes d'accès réseau, processeurs, mémoire [12]

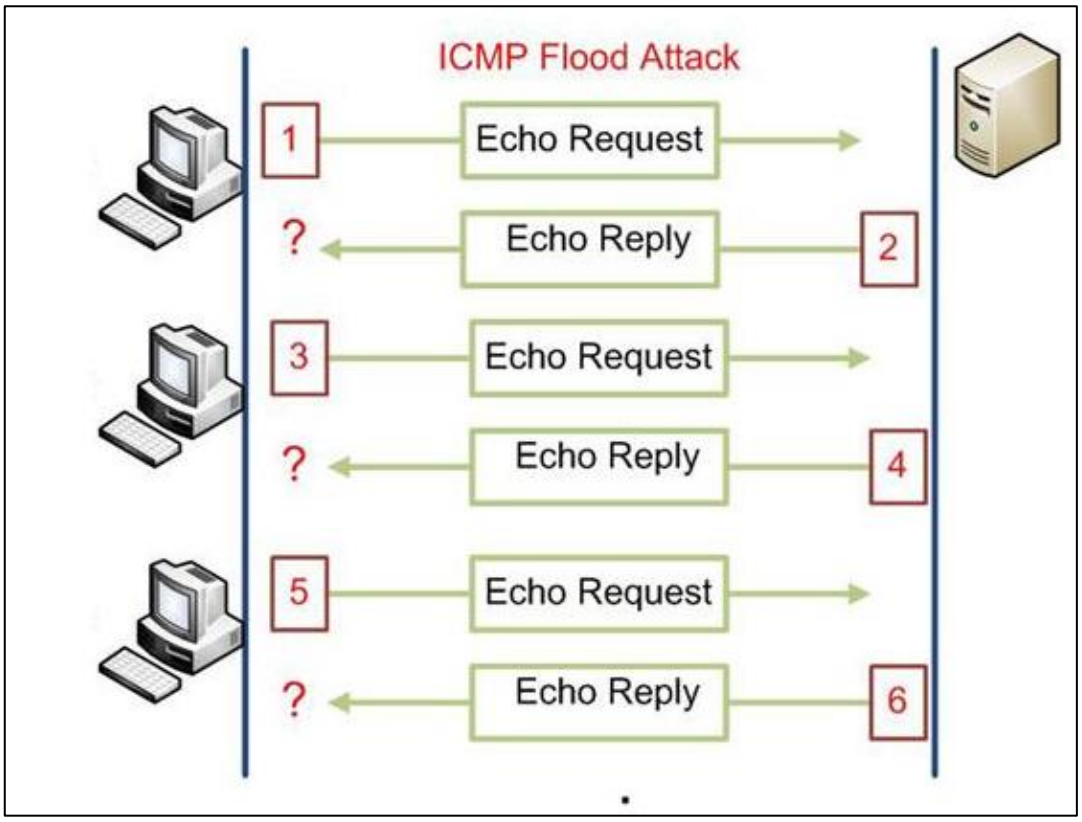

Figure I.7: Attaque DOS de type PING flood.

I.5.Conclusion

IPv6 n'est qu'une nouvelle version du protocole IP. IPv6 respecte donc les principes fondamentaux que l'on connaît de nos jours avec IPv4. Le déploiement de ce nouveau protocole et ses protocoles associés, en particulier l'ICMPv6, a impliqué l'apparition de nouveaux types d'attaques et de différentes natures. Il est vital de mettre en place les mécanismes appropriés pour faire face à ces menaces qui peuvent aboutir à des conséquences dramatiques.

Chapitre II
La théorie des valeurs extrêmes

II.1.Introduction

La théorie des valeurs extrêmes (TVE) est une branche de la statistique dédiée à la modélisation et à l'analyse des événements rares et extrêmes (queues de distribution). Elle évalue la probabilité d'occurrence de valeurs minimales ou maximales dépassant des seuils critiques, contrairement à la statistique classique centrée sur la moyenne. Elle a été développée par Emil Julius Gumbel .

La théorie des valeurs extrêmes a pour objectif essentiel la maîtrise des risques. Ses domaines d'application sont nombreux : on peut citer notamment la finance, l'hydrologie et les domaines informatique et télécommunications.

En télécommunications, la Théorie des Valeurs Extrêmes (TVE) est un outil statistique puissant pour modéliser des événements rares mais à fort impact, dépassant les capacités moyennes de fonctionnement. Elle permet de dimensionner les infrastructures pour éviter les ruptures de service lors de pics extrêmes.

II.2.Distribution limite des maximas

II.2.1.Loi GEV

Le théorème fondamental de la théorie des valeurs extrêmes établit que la limite asymptotique du maximum normalisé converge vers une loi générale, dite loi de Gumbel (au sens large) ou loi des valeurs extrêmes généralisée (GEV) [13]:

$$\frac{M_n - b_n}{a_n} \underset{n\to\infty}{\to} G(x),$$

où

$$G(x) = \exp\left(-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right), 1+\xi\frac{x-\mu}{\sigma} > 0 \qquad \text{II.1}$$

Cette distribution est caractérisée par trois paramètres : μ(paramètre de position), $\sigma > 0$(paramètre d'échelle) et ξ(paramètre de forme), qui contrôle le type de queue de distribution (Gumbel si $\xi = 0$, Fréchet si $\xi > 0$, Weibull si $\xi < 0$).

Selon la valeur du paramètre de forme ξ, la loi GEV se réduit à trois familles distinctes, chacune correspondant à un comportement asymptotique différent des maxima.

Loi de Gumbel ($\boldsymbol{\xi = 0}$)

Par passage à la limite lorsque $\xi \to 0$, on obtient :

$$G(x) = \exp\left(-exp\left(\frac{x-\mu}{\sigma}\right)\right) \qquad \text{II.2}$$

Cette loi se caractérise par une queue exponentielle, c'est-à-dire que la probabilité de dépasser un seuil élevé décroît exponentiellement vite. Elle convient aux événements extrêmes modérés, comme les crues annuelles d'un fleuve, les vents maximaux journaliers ou les rendements financiers modérément extrêmes. Contrairement aux autres cas, la loi de Gumbel ne possède ni queue lourde ($\xi > 0$) ni queue bornée ($\xi < 0$), ce qui en fait un cas intermédiaire.

II.2.2.Loi de Fréchet ($\boldsymbol{\xi > 0}$)

Elle s'écrit :

$$G(x) = \exp\left(\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) \qquad \text{II.3}$$

Sa queue est dite « lourde » (heavy-tail) : la probabilité de dépasser un seuil élevé décroît selon une loi de puissance, plus lentement que l'exponentielle. Cette propriété rend la loi de Fréchet adaptée aux phénomènes très extrêmes dans plusieurs domaine y compris les attaques réseau (débits anormaux), les montants de sinistres en assurance, les précipitations torrentielles, ou encore les rendements financiers lors de krachs boursiers. Plus ξest grand, plus la queue est épaisse et plus les valeurs extrêmes peuvent être démesurées.

II.2.3.Loi de Weibull ($\boldsymbol{\xi < 0}$)

Son expression est formellement identique :

$$G(x) = \exp\left(\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right). \qquad \text{II.4}$$

avec la condition $1 + \xi\frac{x-\mu}{\sigma} > 0$qui impose ici une borne supérieure.

Cette loi possède une queue bornée supérieurement : la variable aléatoire ne peut pas dépasser une certaine valeur maximale théorique. Elle est utile lorsque le phénomène étudié a

une limite naturelle, par exemple la vitesse maximale d'un véhicule sur un tronçon donné, la hauteur maximale des vagues océaniques sous certaines conditions physiques, ou encore la durée de vie maximale d'un composant industriel. Dans ce cas, la distribution des maxima converge vers une loi dont la fonction de répartition atteint 1 en un point fini.

II.3.Méthodes principales en EVT

Il existe plusieurs méthodes d'extraction et de modélisation des valeurs extrêmes, moins courantes que l'approche classique des maxima par bloc (GEV), mais néanmoins utilisées dans des domaines spécifiques tels que la finance, l'environnement ou encore la fiabilité. Nous allons décrire dans ce qui suit les méthodes les plus connues et les plus courantes.

II.3.1.Block Minima (BM)

La méthode traditionnelle GEV modélise les maxima par blocs (Block Maxima). Il existe également une version symétrique : le Block Minima, qui modélise les plus petites valeurs d'un échantillon, utile par exemple pour analyser les temps minimums entre deux pannes, les pertes minimales ou les délais critiques. Le principe consiste à diviser les données en blocs (jours, semaines, mois), à extraire le minimum de chaque bloc, puis à ajuster une distribution GEV. En considérant que $\min(X_1, \dots, X_n) = -\max(-X_1, \dots, -X_n)$, on peut montrer que le minimum normalisé converge vers une loi GEV dont la fonction de répartition s'écrit :

$$G_{\min}(x) = 1 - \exp\left(\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) \qquad \text{II.5}$$

avec la condition $1 + \xi\frac{x-\mu}{\sigma} > 0$.

II.3.2.Méthode des excès en dessous d'un seuil (POT-Lower Tail)

Il s'agit de la version « inverse » de la méthode classique POT (Peaks Over Threshold) basée sur la loi de Pareto généralisée (GPD). Tandis que le POT standard modélise les dépassements au-dessus d'un seuil élevé pour étudier les grandes valeurs extrêmes, la méthode des excès en dessous d'un seuil s'intéresse aux valeurs anormalement petites en modélisant les dépassements en dessous d'un seuil bas. Soit $u$un seuil faible ; on considère les excès négatifs

définis par $Y = u - X$pour $X < u$. En appliquant le théorème de Pickands-Balkema-de Haan, on montre que pour un seuil suffisamment bas, la distribution conditionnelle des excès en dessous du seuil converge vers une loi de Pareto généralisée :

$$P(X < u - y \mid X < u) \approx GPD(y; \xi, \tilde{\sigma}) = 1 - \left(1 + \xi \frac{y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \quad \text{II.6}$$

avec$y > 0$, $1 + \xi y/\tilde{\sigma} > 0$. Dans cette formulation, ξest le paramètre de forme (identique à celui de la queue supérieure si la distribution est symétrique), et $\tilde{\sigma}$un paramètre d'échelle.

II.3.3.Méthode de Régression EVT (EVT-Regression)

Cette approche étend les modèles classiques en permettant aux paramètres de la GEV ou de la GPD de varier en fonction de variables explicatives. Elle permet d'ajuster dynamiquement les probabilités d'événements extrêmes en fonction du contexte opérationnel Ainsi, on considère :

$$\xi(x), \sigma(x), \mu(x)$$

où$x$représente un vecteur de covariables. Ces dépendances peuvent être modélisées via des fonctions linéaires, polynomiales ou splines, par exemple :

$$\mu(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p, \log(\sigma(x)) = \gamma_0 + \gamma_1 x_1 + \cdots + \gamma_p x_p, \xi(x) = \delta_0 + \delta_1 x_1 + \cdots + \delta_p x_p \quad \text{II.7}$$

(avec des contraintes assurant $\sigma(x) > 0$). L'inférence est généralement réalisée par maximum de vraisemblance.

II.3.4.Peaks Over Random Threshold (PORT)

Cette méthode constitue une amélioration du POT (Peaks Over Threshold) classique, dans laquelle le seuil un'est plus fixe mais devient aléatoire. L'idée fondamentale est de choisir un seuil aléatoire qui minimise à la fois la variance, le biais d'estimation et la sensibilité au bruit présent dans les données. Mathématiquement, on définit un seuil aléatoire U(par exemple un quantile empirique mobile ou une fonction d'une variable aléatoire indépendante), et l'on considère les excès au-delà de ce seuil : $\{X_i - U \mid X_i > U\}$. On montre que sous certaines conditions, ces excès suivent encore une loi de Pareto généralisée (GPD). En pratique, le seuil peut être choisi comme un percentile élevé estimé sur une fenêtre glissante, ou comme une réalisation d'une variable aléatoire auxiliaire. Le principal avantage de l'approche PORT est sa

stabilité accrue sur de grands volumes réseau très variables, où un seuil fixe serait soit trop bas (générant trop de faux positifs et un biais important), soit trop haut (conduisant à une variance excessive).

II.3.5.Méthodes Bayésiennes EVT

Dans cette approche, les paramètres de la GPD ou de la GEV (notés $\theta = (\mu, \sigma, \xi)$ou $\theta = (\sigma, \xi)$pour la GPD) sont estimés via l'inférence bayésienne plutôt que par le maximum de vraisemblance classique. On commence par définir une distribution a priori $p(\theta)$sur les paramètres, reflétant les connaissances disponibles avant l'observation des données (par exemple, des contraintes de signe ou des plages raisonnables pour ξ). Ensuite, on actualise cette croyance à l'aide de la vraisemblance $L(\theta; \text{data})$des données extrêmes observées pour obtenir la distribution a posteriori :

$$p(\theta \mid \text{data}) = \frac{L(\theta;\text{data})\, p(\theta)}{\int L(\theta;\text{data})\, p(\theta)\, d\theta} \qquad \text{II.8}$$

L'inférence (moyennes, quantiles, intervalles crédibles) se fait ensuite par échantillonnage de cette distribution a posteriori, typiquement via des méthodes MCMC (Monte Carlo par chaînes de Markov). Les principaux avantages des méthodes bayésiennes en EVT sont : leur capacité à fonctionner avec peu de données extrêmes (situation fréquente en pratique), l'intégration naturelle de l'incertitude des paramètres dans les prédictions, et la production d'intervalles crédibles plus réalistes que les intervalles de confiance asymptotiques.

II.3.6.Quantile Régression for Extrêmes (QR-EVT)

Cette méthode est utilisée pour modéliser directement les quantiles extrêmes (par exemple le quantile à 0,999 ou 99,9 %), sans passer par une estimation complète de la distribution GEV ou GPD. Au lieu d'estimer les paramètres μ, σ, ξpuis d'en déduire les quantiles, QR-EVT modélise le quantile conditionnel d'ordre τ(avec $\tau \to 1$) comme une fonction linéaire ou non linéaire de covariables :

$$Q_{Y|X}(\tau) = \beta_0(\tau) + \beta_1(\tau)X_1 + \cdots + \beta_p(\tau)X_p \qquad \text{II.9}$$

où les coefficients $\beta(\tau)$sont estimés en minimisant une fonction de perte asymétrique adaptée aux queues de distribution. Le principal avantage est que l'on n'a pas besoin de spécifier une forme paramétrique complète de la distribution sous-jacente, ni de recourir à l'estimation classique GEV/GPD.

II.3.7.Méthode des Records (Record Values Theory)

Cette branche de la théorie des valeurs extrêmes s'intéresse non pas à l'ensemble des maxima par blocs, mais aux valeurs record successives, c'est-à-dire aux nouveaux maximums (ou nouveaux minimums) établis au fur et à mesure que l'on observe une séquence de variables aléatoires $X_1, X_2, \dots, X_n$. On dit qu'une observation $X_n$est un record supérieur si $X_n > \max(X_1, \dots, X_{n-1})$. La théorie des records étudie le comportement asymptotique du nombre de records, de l'instant d'apparition des records, et de l'intensité des records. Pour une séquence i.i.d. de densité continue, la probabilité que la n-ième observation soit un record est simplement $1/n$, et le nombre attendu de records après $n$observations est $\sum_{k=1}^{n} 1/k \sim \log n$.

II.4.Estimation des paramètres de loi GEV

IL existe diverses méthodes qui ont été proposées pour estimer les paramètres des Lois GEV. Parmi eux on trouve la méthode basée sur le maximum de vraisemblance (Smith1985 [14]), la méthode des moments pondérés (Greenwoodet al1979 [15]), la méthode des moments (Christopeit1994 [16]),la méthode des L-moments(Hosking1990 [17]).

IL existe également des approches semi-paramétriques pour l'estimationde l'indice de queue. Les plus utilisées en pratique sont l'estimateur de Pickands(Pickands1975 [18]), l'estimateur de Hill(Hill1975 [19]) pour le cas des lois de type Fréch et uniquement etl'estimateur de Dekkers-Einmahl-de Hann(Dekkersetal1989 [20]). Cependant, suivant le type de problématique que l'on a, on va avoir besoinnon seulement d'un estimateur de paramètre de forme, mais aussi d'un estimateur de paramètre d'échelle et estimateur de paramètre de location. C'est typiquement le cas par exemple si l'on s'intéresse à l'estimation d'un quantile extrême.

II.4.1.Estimation paramétrique

L'estimation paramétrique qui considère que les modèles sont connus avec des paramètres inconnus. L'ensemble des valeurs possible pour les paramètres, appelé espace paramétrique se ra noté Θ lequel est inclus dans $\mathbb{R}^d$ où des t la dimension du paramètre θ.

La loi de la variable étudiée est supposée appartenir à une famille de lois pouvant être caractérisée par une forme fonctionnelle connu.

Nous pouvons estimer les paramètres de la GEV par trois méthodes d'estimationsont envisageables: estimation par le Maximum de Vraisemblance (MLE), estimation par les Moments (ME), estimation par les L-Moments (LME).

II.4.1.1.Méthode de maximum de vraisemblance

L'estimation par maximum de vraisemblance donne des résultats asymptotiques efficaces, les estimateurs obtenus convergent sous certaines conditions vers les vraies valeurs des paramètres.

Soient n variables aléatoires $X_1,X_2,...,X_n$ indépendantes identiquement distribuées (iid). On appelle fonction de vraisemblance pour l'échantillon $(X_1,X_2,...,X_n)$ la fonction du paramétre Θ :

$$\mathbf{L(\theta, X\_1, X\_2, \ldots\ldots X\_n)} = \begin{cases} \prod_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{P(X, \theta)} & \textbf{cas discrete} \\ \prod_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{f(X, \theta)} & \textbf{cas continue} \end{cases} \qquad \text{II.10}$$

L'estimation de maximum de vraisemblance de θ est la valeur $\hat{\theta}$ qui rend maximale la fonction de vraisemblece $L(\boldsymbol{\theta}, X_1, X_2, \ldots . X_n)$.

Dans la plupart des cas, la fonction de vraisemblance s'exprime comme un produite, donc $\widehat{\boldsymbol{\theta}}$ sera en générale calculé par maximisation la log-vraisemblece.

Quand $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots\ldots \boldsymbol{\theta}_d)$ et que toutes les dérivées partielles ci-dessous existe, $\widehat{\boldsymbol{\theta}}$ est la solution système d'équation :

$$\begin{cases} \frac{\partial}{\partial \theta_i} \boldsymbol{log\ L(\theta, X_1, X_2 \ldots, X_n) = 0, \forall j \in (1, 2 \ldots)} \\ \frac{\partial^2}{\partial \theta_j^2} \boldsymbol{log\ L(\theta, X_1, X_2 \ldots, X_n \leq 0, \forall j \in (1, 2 \ldots)} \end{cases} \quad \text{II.11}$$

Pour la distribution GEV :

La fonction de log-vraisemblance est donnée par :

Pour $\xi \neq 0$:

$$\boldsymbol{log\ L(\mu, \sigma, \xi) = -n\ log(\sigma) - (\frac{1}{\xi} + 1) \sum_{i=1}^{n} log(1 + \xi \frac{x_i - \mu}{\sigma}) - \sum_{i=1}^{n} (1 + \xi \frac{x_i - \mu}{\sigma})^{\frac{-1}{\xi}}}$$

Pour $\xi = 0$:

$$\boldsymbol{log\ L(\mu, \sigma,) = -n\ log(\sigma) - \sum_{i=1}^{n} (\frac{x_i - \mu}{\sigma}) - \sum_{i=1}^{n} exp\left[-(\frac{x_i - \mu}{\sigma})\right]} \quad \text{II.12}$$

En prenant la dérivée de $log\ L(\mu, \sigma, \xi)$ par rapport àσ, μ et ξ et on obtient le système d'équations suivant :

Pour ξ≠0:

$$\frac{\partial\ log\ L(\mu,\sigma,\xi)}{\partial \mu} = \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^{n} \frac{\xi}{\sigma + \xi(x_i - \mu)} - \frac{1}{\sigma} \sum_{i=1}^{n} (1 + \xi \frac{x_i - \mu}{\sigma})^{\frac{-1}{\xi} - 1} = \quad \text{II.13}$$

$$\frac{\partial\ log\ L(\mu, \sigma, \xi)}{\partial \sigma} = -\frac{n}{\sigma} + (\frac{1}{\xi} + 1) \sum_{i=1}^{n} \frac{\xi(x_i - \mu)}{\sigma^2 + \sigma\xi(x_i - \mu)} - \sum_{i=1}^{n} (\frac{(x_i - \mu)}{\sigma^2})(1 + \xi \frac{x_i - \mu}{\sigma})^{\frac{-1}{\xi} - 1} = 0$$

$$\boldsymbol{\frac{\partial\ log\ L(\mu,\sigma,\xi)}{\partial \xi} = \frac{1}{\xi^2} \sum_{i=1}^{n} log(1 + \xi \frac{x_i - \mu}{\sigma}) - (1 + \frac{1}{\xi}) \sum_{i=1}^{n} \frac{(x_i - \mu)}{\sigma + \xi\ (x_i - \mu)} + \sum_{i=1}^{n} (1 + \xi \frac{x_i - \mu}{\sigma})^{\frac{-1}{\xi}} \left\{ \sum_{i=1}^{n} \frac{(x_i - \mu)}{\sigma\xi + \xi^2 \sigma(x_i - \mu)} - \frac{1}{\xi^2} log(1 + \xi \frac{x_i - \mu}{\sigma}) \right. = 0} \quad \text{II.14}$$

Et pour ξ= 0 :

$$\frac{\partial\ log\ L(\mu, \sigma\ )}{\partial \mu} = 0 = n - \left[\sum exp(-(\frac{x_i - \mu}{\sigma})) - 1\right]$$

$$\frac{\partial \log L(\mu,\sigma)}{\partial\sigma} = 0 = n + \sum \frac{x_i - \mu}{\sigma}\left[exp(-(\frac{x_i - \mu}{\sigma})) - 1\right]$$

Le vecteur$\hat{\theta}(\hat{\mu},\hat{\sigma},\hat{\xi})$ estsolutiondusystèmesuivant:

$$\hat{\boldsymbol{\mu}} = \begin{cases} \frac{\partial \log L(\mu,\sigma,\xi)}{\partial \mu} = 0 \\ \frac{\partial^2 \log L(\mu,\sigma,\xi)}{\partial \mu^2} \leq 0 \end{cases} \quad \text{II.15}$$

$$\hat{\boldsymbol{\sigma}} = \begin{cases} \frac{\partial \log L(\mu,\sigma,\xi)}{\partial \sigma} = 0 \\ \frac{\partial^2 \log L(\mu,\sigma,\xi)}{\partial \sigma^2} \leq 0 \end{cases} \quad \text{II.16}$$

$$\hat{\boldsymbol{\xi}} = \begin{cases} \frac{\partial \log L(\mu,\sigma,\xi)}{\partial \xi} = 0 \\ \frac{\partial^2 \log L(\mu,\sigma,\xi)}{\partial \xi^2} \leq 0 \end{cases} \quad \text{II.17}$$

La résolution de ce système est relativement difficile et n'admet pas généralement de solutions explicites, dans ce cas on fait appel à des méthodes d'optimisation numérique.

II.4.1.2.Méthode de moments

Les moments de population (moyenne, variance, asymétrie, etc.) sont assimilés à leurs moments d'échantillon correspondants, et les paramètres de la distribution sont résolus algébriquement.

La distribution des valeurs extrêmes généralisées (GEV) est une famille de distributions de probabilité continues. La fonction de densité de probabilité de la distribution GEV est définie comme :

Pour$\xi \neq 0$:

$$\boldsymbol{h(x,\mu,\sigma,\xi)} = \frac{\mathbf{1}}{\boldsymbol{\sigma}}\left(\mathbf{1}+\boldsymbol{\xi}\left(\frac{\boldsymbol{x-\mu}}{\boldsymbol{\sigma}}\right)\right)^{-\frac{\mathbf{1}}{\boldsymbol{\xi}}-\mathbf{1}}\boldsymbol{exp}\left[-\left(\mathbf{1}+\boldsymbol{\xi}\left(\frac{\boldsymbol{x-\mu}}{\boldsymbol{\sigma}}\right)\right)^{\frac{-\mathbf{1}}{\boldsymbol{\xi}}}\right] \quad \text{II.18}$$

Pour ξ=0

$$\boldsymbol{h(x,\mu,\sigma,0)} = \frac{\mathbf{1}}{\boldsymbol{\sigma}}\boldsymbol{exp}\left(-\left(\frac{\boldsymbol{x-\mu}}{\boldsymbol{\sigma}}\right)\right)\boldsymbol{exp}\left(-\boldsymbol{exp}\left(-\frac{\boldsymbol{x-\mu}}{\boldsymbol{\sigma}}\right)\right) \quad \text{II.19}$$

tel que:

μ est le paramètre de localisation.

σ est le paramètre d'échelle.

ξ est le paramètre de forme.

Pour ξ≠0:

La moyenne E(X) et la variance Var(X) de la distribution GEV sont données par :

$$\boldsymbol{E(X)} = \boldsymbol{\mu} + \frac{\boldsymbol{\sigma}}{\boldsymbol{\xi}}\left(\boldsymbol{\Gamma}(\mathbf{1}-\boldsymbol{\xi})-\mathbf{1}\right)$$

$$\boldsymbol{Var(X)} = \frac{\boldsymbol{\sigma^2}}{\boldsymbol{\xi^2}}\left[\boldsymbol{\Gamma}(\mathbf{1}-\mathbf{2}\boldsymbol{\xi})-(\boldsymbol{\Gamma}(\mathbf{1}-\boldsymbol{\xi}))^{\mathbf{2}}\right]$$

tel que: Γ(.)estlafonctiongamma.

La moyenne et la variance de la distribution GEV dépendent des paramètres μ, σ et ξ.

Pour trouver la moyenne E(X), nous intégrons le PDF sur toute la plage de x:

$$E(X) = \int_{-\infty}^{+\infty} x h(x)\, dx$$

En branchant l'expression pour h(x), nous obtenons :

$$E(X) = \int_{-\infty}^{+\infty} x\left(\frac{1}{\sigma}\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}-1} exp\left[-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{\frac{-1}{\xi}}\right]\right) dx \quad \text{II.20}$$

Cette intégrale peut être assez complexe et peut ne pas avoir de solution de forme fermée. Par conséquent, des méthodes numériques sont souvent utilisées pour approximer l'intégrale, une approche courante consiste à utiliser des techniques d'intégration numérique telles que la règle de

Simpson, la quadrature gaussienne ou l'intégration de Monte Carlo pour approximer l'intégrale sur toute la plage de x. Ces méthodes consistent à diviser la plage d'intégration en intervalles plus petits et à rapprocher l'intégrale dans chaque intervalle, Une fois l'intégrale approchée, la valeur résultante donne la moyenne (E(X)) de la distribution GEV. Bienque les étapes mathématiques de l'intégration numérique impliquent des techniques informatiques plutôt que des manipulations symboliques, elles fournissent des approximations précises de la moyenne de la distribution GEV.

D'autre part, pour la variance :

$$Var(X) = \int_{-\infty}^{+\infty} x^2 h(x) \qquad \text{II.21}$$

En branchant l'expression pour h(x), nous obtenons:

$$Var(X) = \int_{-\infty}^{+\infty} x^2 \left(\frac{1}{\sigma}\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}-1}\left[-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{\frac{-1}{\xi}}\right]\right) \qquad \text{II.22}$$

Des méthodes numériques sont souvent utilisées pour approximer cette intégrale. Une approche courante consiste à utiliser des techniques d'optimisation numérique pour trouver les valeurs des paramètres qui minimisent la différence entre la moyenne et la variance observées et la moyenne et la variance théoriques de la distribution GEV.

Pour $\xi = 0$:

$$E(X) = \mu + \sigma\gamma \qquad \text{II.23}$$

Où γ est la constante d'Euler-Mascheroni, approximativement égale à 0,5772.

$$Var(X) = \frac{\pi^2}{6}.\sigma^2 \qquad \text{II.24}$$

II.4.1.3.Méthode de L-moments

Les L-moments sont des combinaisons linéaires des statistiques d'ordres. Ont été introduit et extrait de (Silitto 1951 [21]) et revu de manière exhaustive par (Hosking 1990 [22]).

Les L-moments sont analogues aux conventionnels moments mais L-moment ont certains avantages notamment : leur existence, leur unicité et leur robustesse.

L-moment ont trouvés de nombreuses applications dans des domaines de la recherche appliquées tq : le génie civil, la météorologie et hydrolique...

II.4.1.3.1.Représentation de L-moment sous forme de moment de probabilité pondéré

Les moments pondérés en fonction des probabilités (Probability-weighted moment, noté PWM) est une généralisation des moments habituels d'une distribution de probabilité, ont été introduits par (Greenwood et al 1979 [15]). Les moments de probabilité pondérés sont définis par :

$$M_{i,j,k} = E\left[X^i F^j (1 - F)^k\right], i, j, k \in \mathbb{R}$$

Nous pouvons récrire en fonction de quantile, alors :

$$M_{i,j,k} = \int_0^1 \mathbb{Q}(u)^i u^j (1 - u)^k du \qquad \text{II.25}$$

Les deux formes suivantes de PWM sont particulièrement simples et utiles :

$$\begin{cases} \alpha_k = M_{1,0,k} = E[X(1 - F(X))^k] & \text{k est entier positif} \\ \beta_j = M_{1,j,0} = E\left[XF^j(X)\right] & \text{j est entier positif} \end{cases}$$

Donc :

$$\begin{cases} \alpha_k = \int_0^1 \mathbb{Q}(u)(1 - u)^k du \\ \beta_j = \int_0^1 \mathbb{Q}(u)^i u^j du \end{cases} \qquad \text{II.26}$$

Les $\alpha_k$et β_j ont été utilisées comme base de méthode de L-moment pour l'estimation des paramètres de distribution de probabilité car : il existe des relations simples entre les moments de probabilité pondérésα_k ,β_j et les paramètres de distribution de probabilité. Donc : il est possible d'exprimer les paramètres de loi de probabilité comme des fonction PWM. D'autre part, pour calculer les L-moments. On écrit : L-moment comme des combinaisons linéaires de PWM :

$$\lambda_r = (-1)^{r-1} \sum_{k=0}^{r-1} P^*_{r,k} \alpha_k \qquad \text{II.27}$$

Donc :

$$\lambda_1 = \alpha_0 = \beta_0$$

$$\lambda_2 = \alpha_0 - 2\,\alpha_1 = 2\beta_1 - \beta_0$$

$$\lambda_3 = \alpha_0 - 6\,\alpha_1 + 6\alpha_2 = 6\beta_2 - 6\beta_1 + \beta_0$$

$$\lambda_4 = \alpha_0 - 12\,\alpha_1 + 30\alpha_2 - 20\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \qquad \text{II.28}$$

Pour la loi de valeur extrême généralisée (GEV) :

Soit X v.a suit la loi GEV des paramètres, σ, ξ .

Pour trouver les L-moments on utilisons la forme de PWM j , On a :

$$\beta_j = \frac{1}{j+1}\left[\mu - \frac{\sigma}{\xi}\left(1 + \frac{1}{(j+1)^{\xi}}\Gamma(1-\xi)\right)\right], \xi > 1$$

Nous calculonsβ_j, j = 0,…..,3 et remplacent dans λ_r pour trouver :

$$\lambda_1 = \mu - \frac{\sigma}{\xi}(1 + \Gamma(1-\xi))$$

$$\lambda_2 = \frac{\sigma}{\xi}(1 - 2^{\xi})\Gamma(1-\xi)$$

$$\lambda_3 = \frac{\sigma}{\xi}(3 * 2^{-\xi} - 2 * 3^{-\xi} - 1)\Gamma(1-\xi)$$

$$\lambda_4 = \frac{\sigma}{\xi}(1 - 6 * 2^{-\xi} + 10 * 3^{-\xi} - 5 * 4^{-\xi})\Gamma(1-\xi) \qquad \text{II.29}$$

Alors :

par L-moments des paramètres du modèle GEV lorsque $-\frac{1}{2} < \xi < \frac{1}{2}$ sont donnés par :

$$\hat{\xi} \approx 7.8560c + 2.9554c^2$$

$$\hat{\sigma} = \frac{\imath_2\hat{\xi}}{(1 - 2^{-\hat{\xi}})(\Gamma(1-\hat{\xi}))}$$

$$\hat{\mu} = \imath_1 + \frac{\hat{\delta}(1 + \Gamma(1-\hat{\xi}))}{\hat{\xi}}$$

Où

$$c = \frac{2}{3+\tau_3} - \frac{log\,2}{log\,3} \qquad \text{II.30}$$

Et $ı_1$,$ı_2$ sont respectivement les estimateurs des L-moments d'ordre 1 et 2 et du rapport des L-moments Ces estimateurs peuvent être définis à partir des moments de probabilité pondérés :

$$\beta_r = E(XH^r(x))$$

Un estimateur sans biais deβ_r (r > 0) est :

$$b_0 = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{II.31}$$

$$b_r = \frac{1}{n}\sum_{i=1}^{n} \frac{(i-1)(i-2)\ldots(i-r)}{(n-1)(n-2)\ldots(n-33)} X_i \qquad \text{II.32}$$

Et les estimateurs des quatre premiers L-moments sont :

$$ı_1 = b_0$$

$$ı_2 = 2b_1 + b_0$$

$$ı_3 = 6b_2 - 6b_1 + b_0$$

$$ı_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \qquad \text{II.33}$$

En divisant les L-moments d'ordre supérieure à 2 par la mesure de dispersion $ı_2$, on obtient les estimateurs des rapports des L-moments $\tau_r = {}^{ı_r}/_{ı_2}$ avec r=3,4.

Le coefficient de l'asymétrie τ_3 peut s'écrire d'une manière directe de la façon suivante (L-skwenes) est:

$$\tau_3 = \frac{\lambda_3}{\lambda_2} = \sum_{i=1}^{n} \frac{c_i x_i + \frac{\bar{X}}{n}}{ı_2} \qquad \text{II.34}$$

Où

$$c_i = 6\frac{(i-1)(i-2)}{n(n-1)(n-2)} - 6\frac{i-1}{n(n-1)} \qquad \text{II.35}$$

Et $\bar{X}$ est la moyenne empirique de l'échantillon.

II.4.2.Estimation semi-paramétrique

Des procédures statistiques semi-paramétrique appropriées à cette situation, n'exigent pas la connaissance entière de la distribution, mais seulement, sur les queues de distribution. Les estimateurs classiques sont basés sur les plus grandes statistiques d'ordre $X_{n-k,n} \leq \ldots \leq X_{n,n}$ où k est une suite intermédiaire d'entiers liés à la taille de l'échantillon de la façon suivante:

$k = k_{(n)} \longrightarrow \infty$quand$n \to \infty$

II.4.2.1.Estimation de Hill

Cet estimateur a été introduit par (Hill1975 [19]) de forme semi paramétrique, il est basé sur les statistiques d'ordre qui n'est utilisable que pour les distributions appartenant au domaine de Fréchet, l'estimateur de Hill est probablement l'estimateur le plus étudié dans la littérature. Un grand nombre de travaux a été consacré à l'étude de la normalité asymptotique de l'estimateur de Hill, (Daviset Resnick1984[23]) et (HaeusleretTeugels1985[24]) et pour les proprietés de l'estimateur, (Mason1982[25]), (Deheuvels et al 1988[26]).

II.4.2.1.1.Définition

Soit$(k(n))_{n\geq 1}$ une suite d entiers avec 1 < k(n) <n, l'estimateur de Hill est donné par la forme empirique suivante :

$$\hat{\xi}^{(H)}_{(k(n),n)} = \frac{1}{k}\sum_{i=1}^{k} log\, X_{n-i+1,n} - log\, X_{n-k+1,n} \qquad \text{II.36}$$

Preuve On peut justifier la démonstration par la caractérisation :

$$\bar{F}(x) = x^{\frac{-1}{\xi}}\iota(x) \qquad \text{II.37}$$

Avec ι une fonction à variation lente.

Nous avons deux parties : $x^{\frac{-1}{\xi}}$ la partie paramétrique et $\iota(x)$ la partie non paramétrique.

Pour t > 0

$$\lim_{n\to\infty} \frac{\iota(tu)}{\iota(u)} = 1$$

Avec t > 0

$$\lim_{n\to\infty} \frac{\bar{F}(tu)}{\bar{F}(u)} = t^{\frac{-1}{\xi}} \left(\lim_{n\to\infty} \frac{\iota(tu)}{\iota(u)} \right) = t^{\frac{-1}{\xi}} \quad \text{II.38}$$

On déduit l'approximation :

$$\bar{F}(tu) \approx t^{\frac{-1}{\xi}} \bar{F}(u) \quad \text{II.39}$$

Dans l'approximation nous remplaçons tu par x on trouve

$$\bar{F}(x) = \bar{F}(u)\left(\frac{x}{u}\right)^{\frac{-1}{\xi}}$$

La densité f(x) est comme suit :

$$f(x) = \frac{1}{\xi}\frac{\overline{F(u)}}{u}\left(\frac{x}{u}\right)^{\frac{-1}{\xi}-1}$$

L'expression de la fonction de vraisemblance est :

$$L(x_1, x_2 \ldots . x_k; \xi) = \prod_{i=1}^{k} f(x_i)$$

Maximiser la vraisemblance L revient à résoudre le système suivant :

$$\begin{cases} \frac{\partial \log l(x_1, x_2, .. x_k; \xi)}{\partial \xi} = 0 \\ \frac{\partial^2 \log l(x_1, x_2, .. x_k; \xi)}{\partial \xi^2} < 0 \end{cases} \quad \text{II.40}$$

En effet

$$\log l(x_1, x_2, .. x_k; \xi) = \log\left(\prod_{i=1}^{k} f(x_1)\right)$$

$$= \sum_{i=1}^{k} \log f(x_i)$$

$$= \sum_{i=1}^{k} \log\left(\frac{1}{\xi}\frac{\bar{F}(u)}{u}\left(\frac{x_i}{u}\right)^{\frac{-1}{\xi}-1}\right)$$

$$= \sum_{i=1}^{k} \left[\log \bar{F}(u) - \log(\xi u) - \left(\frac{1}{\xi} + 1\right) \log\left(\frac{x_i}{u}\right)\right] \quad \text{II.41}$$

Résoudre le système revient à résoudre l'équation suivant :

$$\sum_{i=1}^{k}\left[\frac{-u}{\xi u}+\frac{1}{\xi^2}log(\frac{x_i}{u})\right]=0$$

$$-k\xi\ +\ \sum_{i=1}^{k} log\frac{x_i}{u}=\ 0$$

La solution de cette équation est de la forme :

$$\hat{\xi}^{(H)}=\ \frac{1}{k}\sum_{i=1}^{k} log\frac{x_i}{u}=\frac{1}{k}\sum_{i=1}^{k}[log\ x_i-\ log\ u] \qquad \text{II.42}$$

On remplace x_i et u par leurs équivalents dans la statistique d'ordre on trouve la formule de l'estimateur de Hill :

$$\hat{\xi}^{(H)}_{(k(n),n)}=\frac{1}{k}\sum_{i=1}^{k}\left[log\ X_{n-i+1,n}-log\ X_{n-k+1,n}\right] \qquad \text{II.43}$$

II.4.2.1.2.Propriétés de l'estimateur de Hill

1**. Convergence en probabilité**

Soit$X_1,\ldots\ldots,X_n$un échantillon de variables aléatoires ii.d de fonction de réparation F ∈MDA(H_ξ) , $\xi>0$.

Soit $(k(n))_{n>1}$ une suite d'entiers telle que 1 < k(n) <n , si

$$\frac{k(n)}{n}\ \to 0\ etk(n)\to\infty$$

Alors

$$\hat{\xi}^{(H)}_{((k(n)),n)}\xrightarrow{P}\xi\ \ ,n\to\infty$$

Convergence forte (convergence presque sûre) Si

$$\frac{k(n)}{n}\to 0\ \frac{k(n)}{log\ log\ n}\ \to\infty\ \ ,n\to$$

∞ II.44

Alors

$$\hat{\xi}^{(H)}_{((k(n)),n)} \overset{Ps}{\to} \xi \ , n \to \infty$$

2. Normalité asymptotique

Soit$(k(n))_{n>1}$ une suite d'entiers telle que 1 < k(n) <n , si$\frac{k(n)}{n} \to 0$ $etk(n) \to \infty$,si la condition de l'ordre est satisfaite avec[19] :

$$k(n)A(\frac{n}{k(n)}) \to 0 \ quandn \to \infty$$

$$k(n)^{\frac{1}{2}}(\hat{\xi}^{(H)}_{(k(n),n)} - \xi \ ) \to N(0, \xi^2$$

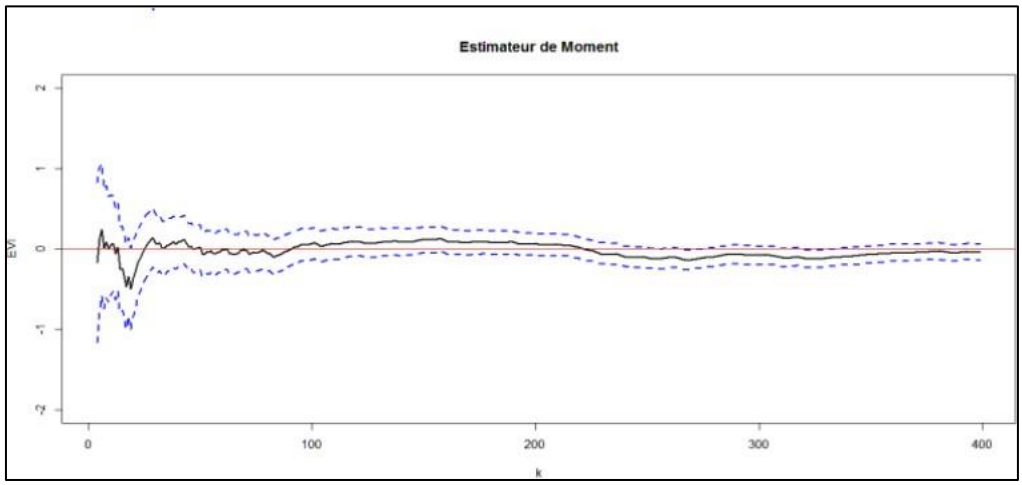

Figure II.1: Représentation graphique de l'estimation de Hill

II.4.2.2.Estimation de Pickands

L'estimateur de Pickands est construit en utilisant les statistiques d'ordres, il a l'avantage d'être valable quel que soit le domaine d'attraction de la distribution et par conséquent du domaine de définition de l'indice des valeurs extrêmes.

Cet estimateur a été introduit par (Pickands et al 1975[18]) et revisité (Drees 1996[27]), (Drees and Kaufmann 1998[28]) d'une part, (Dekkers et al 1989[29]) d'autre part, ils ont établi sa consistance faible ainsi que sa normalité asymptotique.

II.4.2.2.1.Définition

On suppose que $\{X_i, i = 1, \ldots, n\}$ une suite de variables aléatoires indépendantes de loi F appartenant à l'un des domaines d'attractions.

Soit $(k(n))_{n\geq 1}$une suite d'entiers et 1 <kn< n un estimateur de Pickands est défini par :

$$\hat{\xi}^{(p)}_{(k(n),n)} = \frac{1}{log\,2} log(\frac{X_{n-k+1,n}-X_{n-2k+1,n}}{X_{n-2k+1,n}-X_{n-4k+1,n}}) \quad \text{II.45}$$

II.4.2.2.2.Propriétés de l'estimateur de Pickands

1. Convergence en probabilité

Soit$x_1,\dots,X_n$un échantillon de variables aléatoires i.i.d de fonction de répartition $\in M\ \mathrm{D}\ A\ (H_\xi)$, si

$$\boldsymbol{n \to \infty\ , k(n) \to \infty\ et\ \frac{k(n)}{n} \to \infty} \quad \text{II.46}$$

alors

$$\boldsymbol{\hat{\xi}^{(p)}_{(k(n),n)} \xrightarrow{p} \xi}$$

Convergence forte (convergence presque sûre) [18].

Si

$$\boldsymbol{\frac{k}{n} \to \infty\ et \frac{k(n)}{log\ log(n)} \to \infty\ , n \to \infty}$$

Alors

$$\boldsymbol{\hat{\xi}^{(p)}_{(k(n),n)} \xrightarrow{p.s} \xi\ , n \to \infty}$$

2. Normalité asymptotique

Sous les conditions additionnelles sur la suite k(n) et la fonction de répartition F on a[18] :

:

$$\boldsymbol{k(\hat{\xi}^{(p)}_{(k(n),n)} - \xi) \to N(0, v(\xi))}$$

Avec

$$\boldsymbol{k(n)A(\frac{n}{k(n)}) \to 0 quand n \to \infty}$$

ou

$$\boldsymbol{v(\xi)} = \frac{\xi^2(2^{2\xi+1}+1)}{(2(2^{\xi}-1)\,log\,2)^2} \qquad \text{II.47}$$

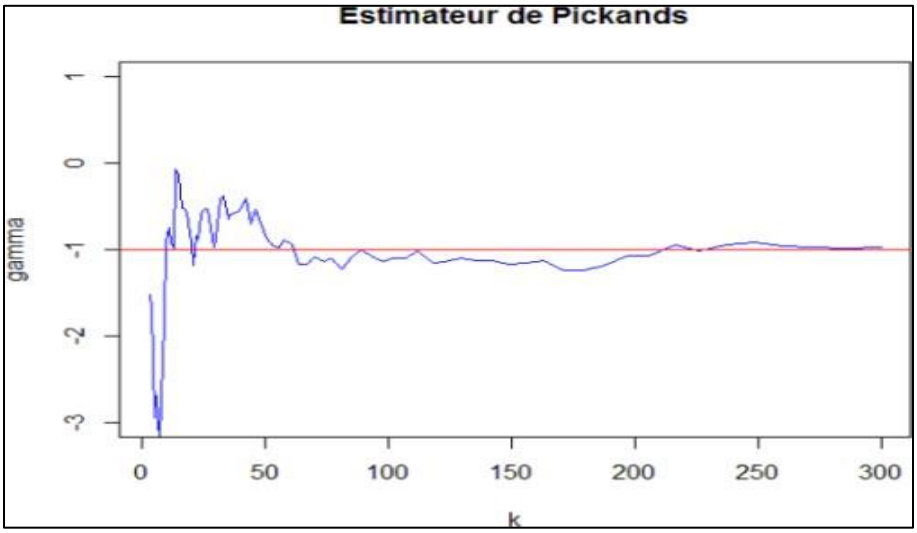

Figure II.2:Représentation graphique de l'estimateur de Pickands

II.4.2.3.Estimateur du moment

Cet estimateur introduit par (Dekkersetal1989[20]) est une généralisation de l'estimateur de Hill, valable pour tout $\xi \in \mathbb{R}$:

$$\hat{\boldsymbol{\xi}}^{(\boldsymbol{M})}_{(\boldsymbol{k(n)},\boldsymbol{n})} = \hat{\boldsymbol{\xi}}^{(\boldsymbol{H})}_{(\boldsymbol{k(n)},\boldsymbol{n})} + \mathbf{1} - \frac{\mathbf{1}}{\mathbf{2}}\left(\mathbf{1} - \frac{\hat{\boldsymbol{\xi}}^{(\boldsymbol{H})}_{(\boldsymbol{k(n)},\boldsymbol{n})}}{\boldsymbol{M}^{(\mathbf{2})}_{\boldsymbol{n}}}\right) - \mathbf{1}$$

Où

$$\boldsymbol{M}^{(\boldsymbol{r})}_{\boldsymbol{n}} := \frac{\mathbf{1}}{\boldsymbol{k}}\sum_{\boldsymbol{i}=\mathbf{0}}^{\boldsymbol{k}} (\boldsymbol{log}\, \boldsymbol{X}_{\boldsymbol{n-i+1,n}} - \boldsymbol{log}\, \boldsymbol{X}_{\boldsymbol{n-k+1,n}})^{\boldsymbol{r}}\,; \boldsymbol{r} = \mathbf{1}, \mathbf{2} \qquad \text{II.48}$$

Propriétés asymptotiques de $\hat{\xi}^{(M)}_{(k(n),n)}$[20]

Supposons que $F \in\ D(H_{\xi}), \xi \in \mathbb{R}\,, k \to \infty\ et \frac{k}{n} \to 0\ quand n \to \infty$:

Consistance faible

$$\hat{\boldsymbol{\xi}}^{(\boldsymbol{M})}_{(\boldsymbol{k(n)},\boldsymbol{n})} \xrightarrow{\boldsymbol{p}} \boldsymbol{\xi}\,, \boldsymbol{quand n} \to \infty$$

Consistance forte

$$\hat{\xi}^{(M)}_{(k(n),n)} \xrightarrow{p.8} \xi \, , quandn \to \infty$$

Normalité asymptotique pour $\xi \geq 0$:

$$\sqrt{k}\,(\hat{\xi}^{(M)}_{(k(n),n)} - \xi) \xrightarrow{d} N(0, \eta^2) \, , quandn \to \infty$$

tel que :

$$\eta^2 = \begin{cases} 1 + \xi^2 & , \xi \geq 0 \\ (1 - \xi^2)(1 - 2\xi)(4 - \frac{1-2\xi}{1-3\xi} + \frac{(5-11\xi)(1-2\xi)}{(1-3\xi)(1-4\xi)}) & , \xi < 0 \end{cases} \quad \text{II.49}$$

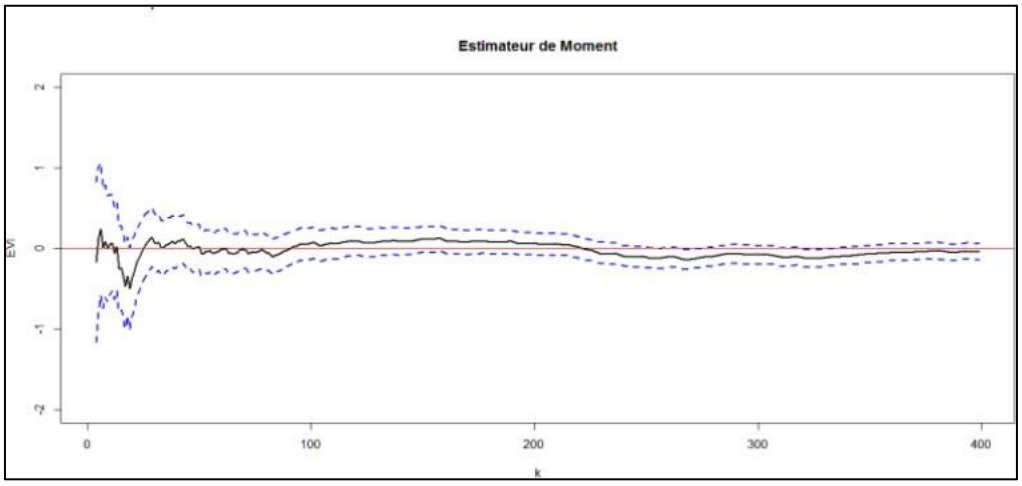

Figure II.3: Représentation graphique de l'estimateur de Moment.

II.4.2.4.Le choix du nombre k

La difficulté consiste à choisir le nombre k de statistiques d'ordre extrême à utiliser. Les résultats concernant les estimateurs de l'indice des valeurs extrêmes sont asymptotiques : ils sont obtenus lorsque$k \to \infty \, et \frac{k}{n} \to 0$. Comme en pratique, on ne dispose que d'un nombre d'observations n fini, il s'agit de choisir k de manière à ce que l'on dispose de suffisamment de matériel statistique tout en restant dans la queue de distribution (k < n).

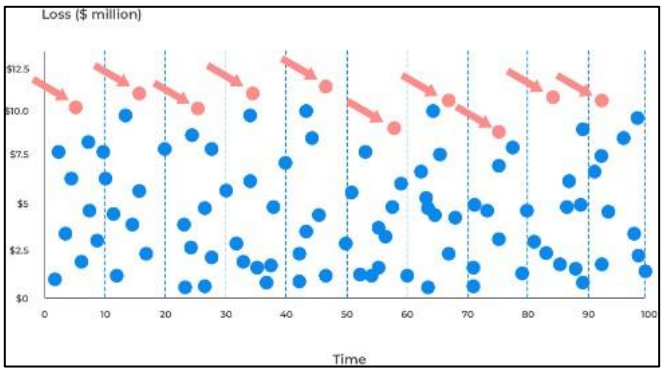

Figure II.4: Diagramme schématique de la méthode des blocs maxima

II.4.2.4.1.Méthode graphique

C'est une méthode la plus simple pour la détermination de k. Elle consiste à tracer le graphe $(k, \hat{\xi}_{(k(n),n)})$ et de prendre la valeur où $(k, \hat{\xi}_{(k(n),n)})$ devient horizontal. Cet estimateur est valable seulement dans le domaine d'attraction de Fréchet c'est-à- dire si $\xi > 0$.

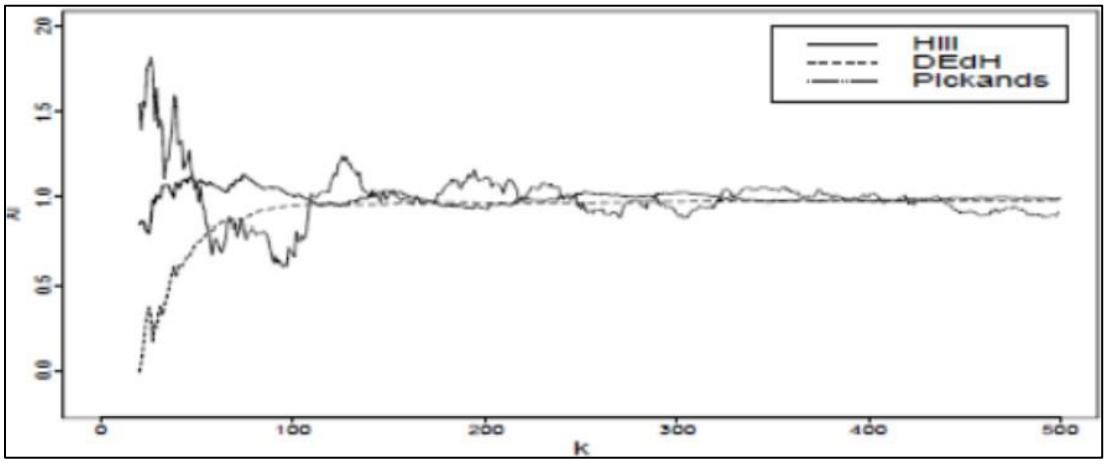

Figure II.5: Le graphe de $(\boldsymbol{k}, \hat{\boldsymbol{\xi}}_{\boldsymbol{n}})$ pour les différents estimateurs de $\boldsymbol{\xi}$

II.4.2.4.2.Méthode analytique

Il est nécessaire pour donner une précision à l'estimateur $\hat{\xi}_{(k(n),n)}$, de calculer l'erreur quadratique moyenne (MSE), elle est en fonction de k :

$$\boldsymbol{MSE}(\hat{\boldsymbol{\xi}}_{(\boldsymbol{k(n)},\boldsymbol{n})}) = \boldsymbol{E}(\hat{\boldsymbol{\xi}}_{(\boldsymbol{k(n)},\boldsymbol{n})} - \boldsymbol{\xi})^{\boldsymbol{2}}$$

$$= \boldsymbol{B}_{\boldsymbol{iais}^{\boldsymbol{2}}}(\hat{\boldsymbol{\xi}}_{(\boldsymbol{k(n)},\boldsymbol{n})}) + \boldsymbol{Var}(\hat{\boldsymbol{\xi}}_{(\boldsymbol{k(n)},\boldsymbol{n})}) \qquad \text{II.50}$$

Le choix optimal de k , correspond à minimiser la MSE.

II.4.2.5.Estimation des quantiles extrêmes

Estimation des quantiles extrêmes joue un rôle important dans le contexte de la gestion des risques où il est crucial d'évaluer de manière adéquate le risque d'une grande perte qui se produit très rarement. On considère$X_1, X_2, \ldots . X_n$laréalisation de n variables réelles indépendantes et de fonction de répartition commune F supposée continue. Pour 0 < p < 1, le quantile d'ordre (1-p) de la fonction de distribution F noté $x_p$est défini comme étant la solution de l'équation :

$$1 - \mathrm{F}(x) = p$$

On dé nit les quantiles extrêmes par :

$$x_{1-p} = F^{\leftarrow}(1-p) = \mathbb{Q}(1-p)\ quandp \to 0$$

On veut estimer le quantile d'ordre p où p est strictement inférieur à 1/n. L'estimateur de quantile extrême obtenu par la méthode des blocks s'écrit sous la forme :

$$\widehat{\boldsymbol{x}}_{\boldsymbol{p}} = \begin{cases} \widehat{\boldsymbol{\mu}} - \frac{\widehat{\boldsymbol{\sigma}}}{\widehat{\boldsymbol{\xi}}}\left[\mathbf{1} - (-\boldsymbol{log}(\mathbf{1}-\boldsymbol{p}))^{-\widehat{\boldsymbol{\xi}}}\right] & ,\boldsymbol{\xi} \neq \mathbf{0} \\ \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\sigma}}\,\boldsymbol{log}(-\boldsymbol{log}(\mathbf{1}-\boldsymbol{p})) & ,\boldsymbol{\xi} = \mathbf{0} \end{cases} \qquad \text{II.51}$$

Où$\hat{\mu}, \hat{\sigma}, \hat{\xi}$et sont estimateurs des paramètres de la loi GEV . Supposons maintenant que nous disposions d'un quantile x_p et que nous désirions estimer la période de retour T, associée à la probabilité d'excès au-delà de x_p. Cette période s interprète comme une moyenne d'unités de temps séparant un événement de grandeur donnée x_pd'un second événement d'une grandeur égale ou supérieure.

II.4.2.6.Estimation de période de retour

Nous définissons le niveau de retour comme la valeur x_τ , telle que nous espérons détecter en moyenne un seul dépassement de cette quantité au bout de T périodes c'est-à-dire :

$$E\left(\sum_{i=1}^{T} 1_{x_i > x_T}\right) = 1$$

$$\leftrightarrow P(X_i > X_T) = \frac{1}{T}$$

$$\leftrightarrow 1 - F(X_T) = \frac{1}{T}$$

L'estimateur d'un niveau de retour d'ordre T revient à l'estimation d'un quantile extrême d'ordre $p = 1 - \frac{1}{T}$:

Un estimateur d'ordre $\hat{p}^{GEV}$ du quantile x_p pour la loi des GEV est donné par :

$$\hat{\mathrm{p}}^{\mathrm{GEV}} = \begin{cases} \exp\left(-\left(1 + \hat{\xi}\left(\frac{\mathrm{x_p} - \hat{\mu}}{\hat{\sigma}}\right)\right)^{\frac{-1}{\hat{\xi}}}\right) & \text{si } \hat{\xi} \neq 0 \\ \exp\left(-\exp - \left(\frac{\mathrm{x_p} - \hat{\mu}}{\hat{\sigma}}\right)\right) & \text{si } \hat{\xi} = 0 \end{cases} \qquad \text{II.52}$$

D'où :

$$\hat{\mathrm{T}}^{\mathrm{GEV}} = \begin{cases} \left(\left(1 - \exp\left(-\left(1 + \hat{\xi}\left(\frac{\mathrm{x_p} - \hat{\mu}}{\hat{\sigma}}\right)\right)^{\frac{-1}{\hat{\xi}\hat{\xi}}}\right)\right)^{-1} & \text{si } \hat{\xi} \neq 0 \\ \left(\left(1 - \exp\left(-\exp - \left(\frac{\mathrm{x_p} - \hat{\mu}}{\hat{\sigma}}\right)\right)\right)\right)^{-1} & \text{si } \hat{\xi} = 0 \end{cases} \qquad \text{II.53}$$

$$\hat{\mathrm{T}}^{\mathrm{GEV}} = \frac{1}{1 - \mathrm{H}_{\hat{\xi}}(\mathrm{x_p})} \qquad \forall \hat{\xi} \in \mathbb{R}$$

II.5.Conclusion

Dans ce chapitre, nous avons présenté les fondements théoriques de la théorie des valeurs extrêmes (EVT) ainsi que ses principales approches, notamment les méthodes GEV et POT-GPD. Ces modèles permettent de caractériser le comportement des événements rares en se concentrant sur les queues de distribution, ce qui est particulièrement adapté à l'analyse des anomalies dans les réseaux informatiques.

Nous avons également mis en évidence l'importance du choix des paramètres, tels que le seuil pour la méthode POT et les quantiles pour la méthode GEV, qui influencent directement la qualité de la détection. Enfin, ce chapitre a permis de montrer que l'EVT constitue un cadre statistique solide pour la modélisation des extrêmes dans différents domaines, notamment la cybersécurité. Dans le chapitre suivant, nous explorerons l'utilité de ces méthodes statistiques extrêmes pour la détection des cyberattaques, en les appliquant spécifiquement aux réseaux IPv6.

Chapitre III
Simulations et interprétations

III.1.Introduction

Une anomalie réseau peut être définie comme une activité ou un événement qui ne correspond pas au comportement de trafic attendu ou établi. Elles peuvent indiquer une variété d'événements tels que des défaillances techniques (panne d'un équipement réseau), des congestions de trafic, des erreurs opérationnelles (mauvaise configuration par exemple) ou des variations de l'activité. Cela inclut également le trafic dû à des cyberattaques (intrusions, malwares, phishing, DDoS, exfiltration de données,. . .). A mesure que les réseaux deviennent plus complexes et que les attaques deviennent plus sophistiquées, les techniques de détection doivent constamment être améliorées.

C'est dans cette optique que nous présentons, dans ce chapitre, une exploration approfondie des performances des méthodes basées sur la théorie des valeurs extrêmes EVT appliquée à la détection des cyberattaques dans les réseaux IPv6. L'objectif principal de ce chapitre sera d'évaluer les performances des méthodes GEV et POT-GPD sous différents scénarios et conditions de cyberattaques (SYN flood, UDP flood, NAS,…). Nous confronterons ces deux approches à travers une étude comparative rigoureuse, en analysant leur efficacité respective pour la détection des attaques. Cette évaluation sera réalisée en utilisant différents scénarios d'attaques et en s'appuyant sur plusieurs bases de données représentatives des environnements IPv6.

III.2.Méthodologie EVT pour la détection des cyberattaques

La figure III.1 récapitule la méthode d'utilisation des méthodes EVT pour une détection des attaques dans un réseau IPv6.

III.2.1.Définir les indicateurs de sécurité à surveiller

Il s'agit de choisir les variables quantitatives susceptibles de présenter des valeurs extrêmes en cas d'activité malveillante. Parmi les indicateurs couramment qu'on peut utiliser on trouve : le volume de trafic réseau par unité de temps, le nombre de connexions par adresse IP, la taille des paquets, le taux d'erreurs ou de requêtes rejetées, le temps de réponse des serveurs, les débits CPU, RAM ou disque lors d'une intrusion, ainsi que le nombre de tentatives de connexion (typique d'une attaque par force brute). Ces variables doivent être mesurées en continu pour alimenter le modèle.

III.2.2.Collecter et pré-traiter les données

Avant d'appliquer les modèles EVT, une étape de prétraitement est indispensable. Elle consiste à nettoyer les données en corrigeant ou supprimant les valeurs manquantes et les doublons, à dédupliquer les flux réseau pour éviter les redondances, à normaliser les variables si leurs échelles diffèrent fortement, et éventuellement à séparer les données par période (jour/nuit, semaine/week-end) pour tenir compte des cycles saisonniers. L'objectif final est d'obtenir une série temporelle propre et stationnaire par morceaux.

III.2.3.Choisir l'approche EVT

Deux approches principales s'offrent à l'analyste. L'approche dite « Block Maxima » consiste à découper les données en blocs de temps égaux (par exemple 1 heure ou 1 jour), à extraire la valeur maximale dans chaque bloc, puis à modéliser ces maxima par une loi GEV. Cette méthode est utile pour une surveillance périodique, comme le trafic maximal par heure.

L'approche POT fixe un seuil élevé $u$et étudie toutes les observations qui dépassent ce seuil, en les modélisant par une loi GPD. POT est plus fine, plus efficace en termes d'utilisation des données extrêmes, et mieux adaptée à la détection d'intrusion en temps réel.

III.2.4.Ajuster le modèle EVT

Une fois la méthode choisie, on procède à l'ajustement du modèle. Pour l'approche POT, on sélectionne un seuil u(typiquement le quantile 95 % ou 99 % de la distribution empirique), on extrait les excès $X_u = X - u$pour les observations dépassant u, puis on ajuste une distribution de ParetoGPD de paramètres ξ(forme) et $\tilde{\sigma}$(échelle). Pour l'approche Block Maxima, on estime directement les paramètres de la loi GEV : μ(position), $\sigma > 0$(échelle) et ξ(forme). L'estimation des paramètres peut être réalisée par la méthode du maximum de vraisemblance MLE, par des méthodes bayésiennes, ou encore par la méthode des moments.

III.2.5.Définir un seuil d'alerte « extrême »

A partir du modèle EVT ajusté (GEV ou GPD), on calcule des quantiles extrêmes, par exemple le quantile à 99,9 % ou 99,99 %. On détermine également la probabilité qu'une nouvelle observation dépasse une valeur donnée selon le modèle. Un seuil d'alerte est alors établi : toute observation dont la probabilité d'occurrence sous le modèle ajusté est très faible (typiquement inférieure à 0,1 % ou 0,01 %) est considérée comme potentiellement malveillante. Ce seuil peut être défini comme le quantile extrême correspondant à une probabilité de dépassement choisie.

III.2.6.Détecter les attaques en temps réel

Chaque nouvelle observation x_t(par exemple le trafic à l'instant t) est comparée au seuil d'alerte déterminé à l'étape précédente. Si $x_t$est inférieur ou égal au seuil, l'observation est considérée comme normale. En revanche, si $x_t$dépasse le seuil, une alerte potentielle est déclenchée. Cette étape fournit un mécanisme d'alerte automatique, totalement indépendant des signatures d'attaque connues, ce qui permet de détecter des attaques zero-day ou des comportements anormaux inédits.

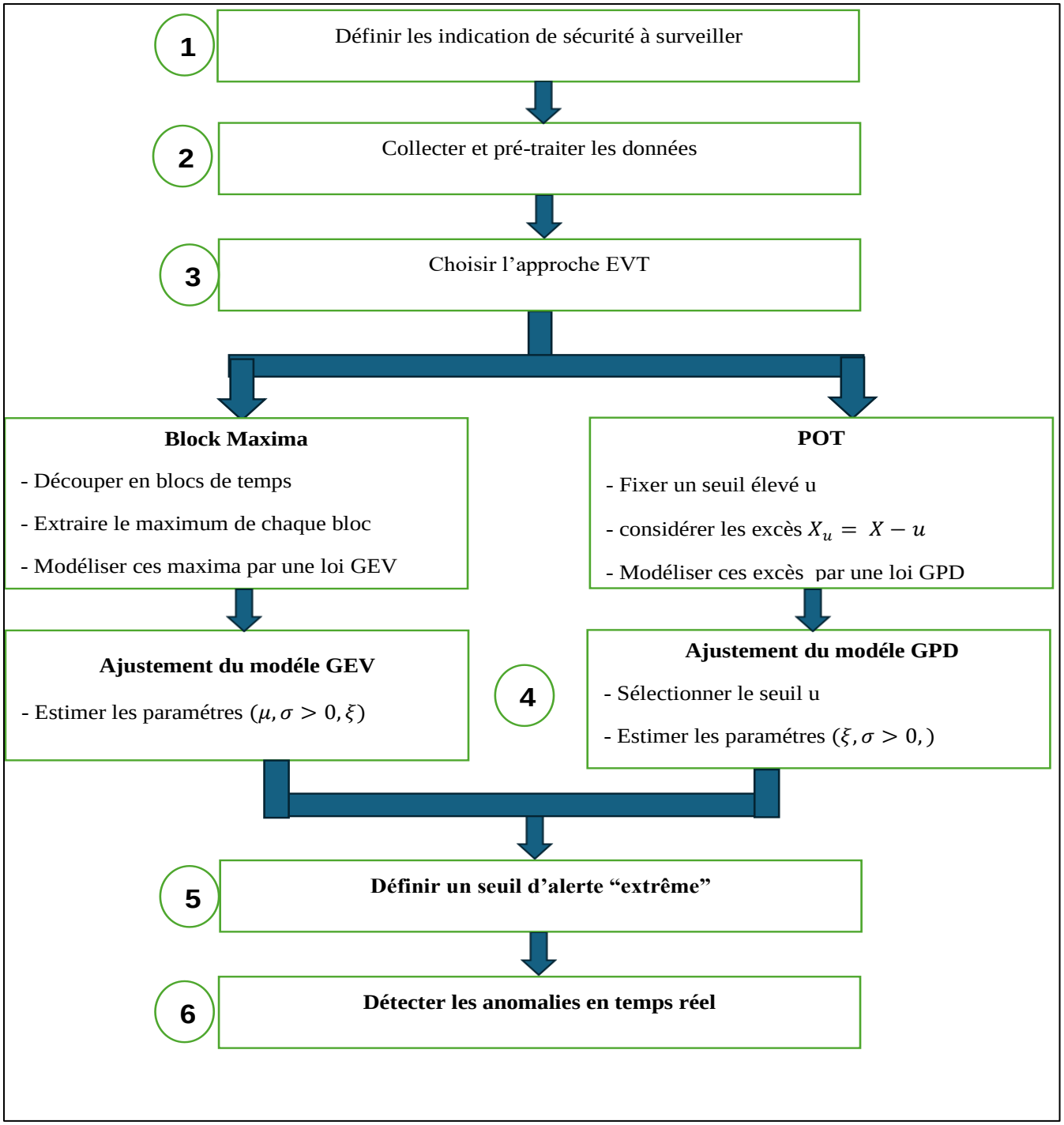

Figure III.1 :Méthodologie de détection des cyber-attaques par EVT

III.3.Description des Datasets utilisées :

L'ensemble de données comprend les caractéristiques du trafic UDP, ICMP et TCP, malveillant, provenant d'un réseau émulé SDN. Ces caractéristiques ont été extraites et calculées à partir des statistiques du commutateur. Cet ensemble de données peut être utilisé pour la classification du trafic et la détection des attaques DDoS dans un environnement SDN, à base de protocoles IPv6.[30]

III.3.1.Etapes de reproduction

Dans un premier temps, Mininet a été utilisé pour configurer l'environnement SDN et créer les topologies réseau. Le contrôleur RYU a ensuite été connecté à plusieurs commutateurs. Trois topologies différentes ont été mises en œuvre. Une fois les topologies réseau initialisées, l'étape suivante consistait à générer du trafic. Le trafic d'attaque a été généré à l'aide de Hping3, tandis que le trafic normal a été généré à l'aide d'Iperf et de l'outil ping. Les protocoles TCP, UDP et ICMP ont été utilisés pour le trafic malveillant et le trafic normal, respectivement. Après la génération du trafic, une requête de statistiques a été envoyée au commutateur via le contrôleur SDN à l'aide du gestionnaire OFPFLOWSTATREQUEST, et une réponse a été reçue à l'aide du gestionnaire OFPFLOWSTATREPLY. La méthode OFPFLOWSTATREQUEST permet de demander au commutateur des informations statistiques telles que l'adresse source, l'adresse de destination, les flux, les paquets de flux, la durée des flux, etc. Afin de recevoir une réponse à cette requête, le gestionnaire OFPFLOWSTATREPLY est appelé pour récupérer les statistiques de flux. Un script basé sur Python a été développé pour collecter et calculer les statistiques des commutateurs dans un intervalle de temps prédéterminé, qui ont été enregistrées dans un fichier CSV. La base de données contient un trafic IPv6 avec des attaques, entre autres, de type : attaques ICMP Flood, attaques à base de TCP, UDP Flood, Port Scan IPv6.

III.4.Résultats et interprétations

III.5.Détection les attaques par POT-GPD et GEV

III.5.1.1.Détection les attaques ICMP-FLOOD

Les figures III.2, III.3 et III.4 présentent les résultats de détection des attaques ICMP-Flood dans un environnement IPv6 à l'aide des méthodes EVT POT-GPD et GEV. Les résultats montrent que les deux approches permettent d'identifier les comportements anormaux du trafic réseau et de distinguer efficacement le trafic normal du trafic malveillant. Pour les faibles valeurs de quantile, notamment à 0,90, les seuils de détection restent relativement bas, ce qui rend les deux méthodes plus sensibles aux variations du trafic et entraîne la détection d'un grand nombre d'anomalies. Lorsqu'il augmente à 0,95, les seuils deviennent plus élevés et les anomalies détectées correspondent principalement aux pics importants générés par les attaques ICMP-Flood dans la partie centrale du trafic réseau. Enfin, pour un quantile de 0,99, les seuils deviennent très stricts, particulièrement pour la méthode GEV qui ne détecte plus aucune anomalie, alors que la méthode POT-GPD conserve encore une capacité de détection des pics les plus significatifs. Ces résultats montrent ainsi que la méthode POT-GPD offre une meilleure sensibilité et une détection plus efficace des attaques ICMP-Flood dans les réseaux IPv6, tandis que la méthode GEV est davantage adaptée à l'identification des événements extrêmes les plus critiques.

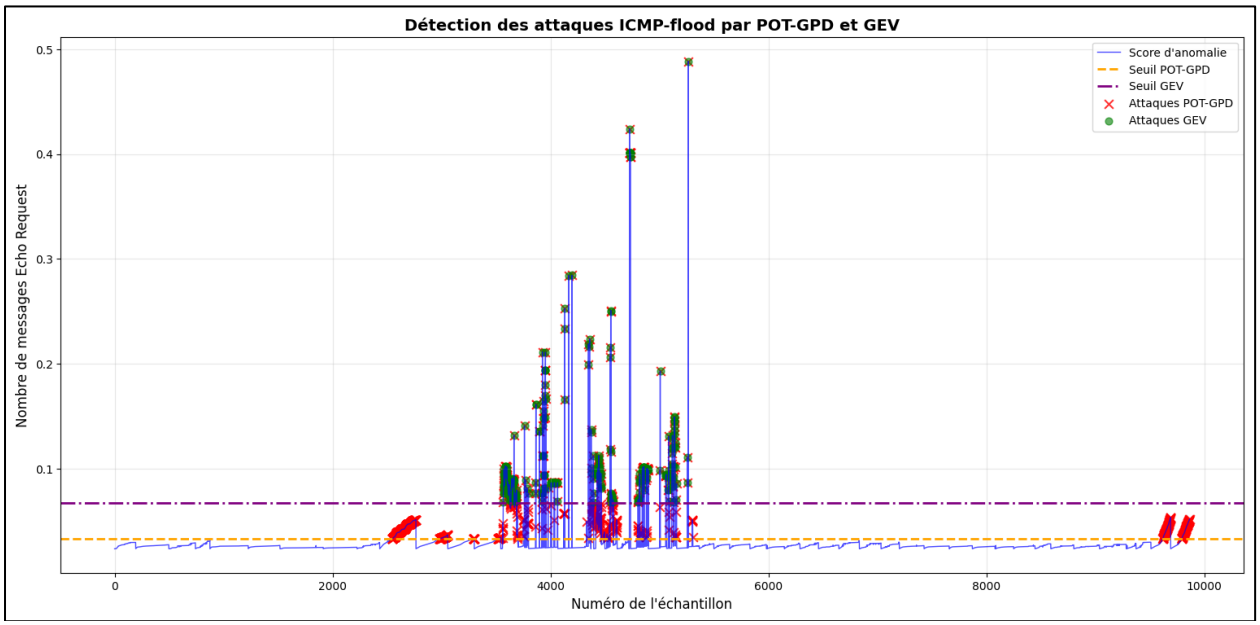

Figure III.2 : Détection des attaques ICMP flood par les méthodes POT-GPD et GVE (Quantile = 0,90)

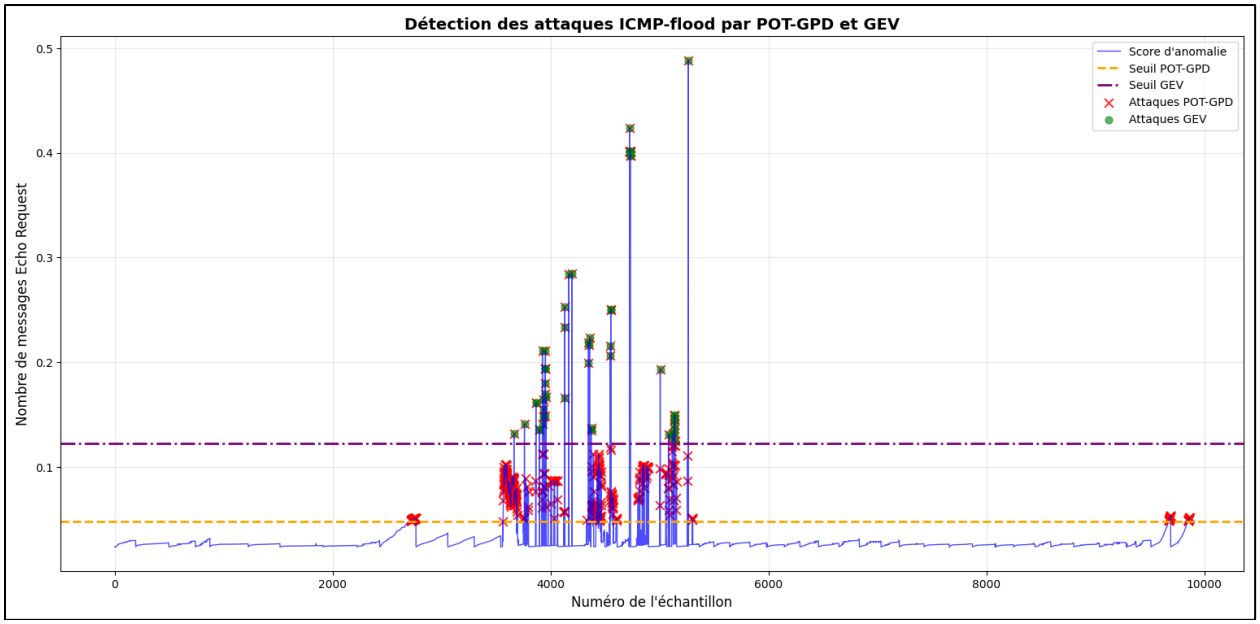

Figure III.3 : Détection des attaques ICMP flood par les méthodes POT-GPD et GVE (Quantile= 0,95)

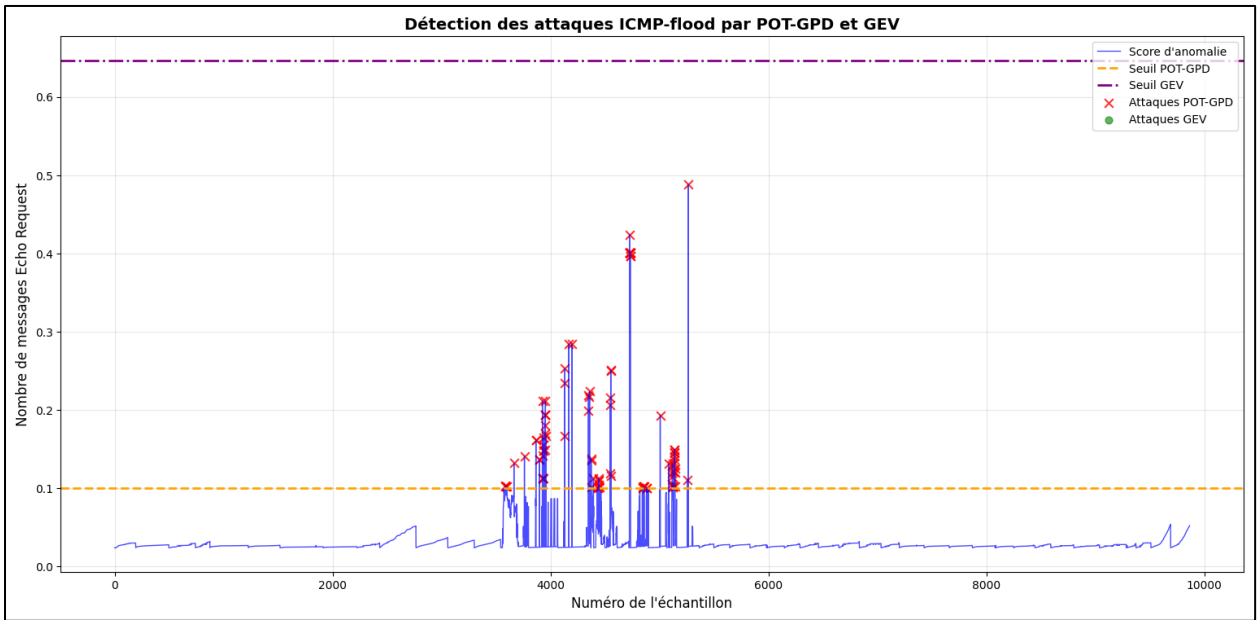

Figure III.4 : Détection des attaques ICMP flood par les méthodes POT-GPD et GVE (Quantile=0,99)

III.5.1.2.Comparaison de l'évolution des seuils EVT et du nombre d'attaques

La figure III.5 reporte la comparaison des deux méthodes EVT en termes d'évolution du seuil de détection ainsi que le nombre des attaques ICMP flood détectées. Les courbes à gauche présente l'évolution des seuils de détection en fonction des différentes valeurs du quantile. Ils montrent comment les seuils calculés par les méthodes EVT varient selon le niveau de sensibilité choisi pour la détection des anomalies. Quant aux barres de droite, ils représentent le nombre d'attaques détectées pour chaque valeur du quantile.

Quant à l'évolution des seuils de détection calculés par les deux méthodes EVT (GEV et POT-GPD) en fonction des valeurs du quantile variant de 0,90 à 0,99, on observe que le seuil de la méthode POT-GPD augmente progressivement de 0,08 (quantile 0,90) à 0,16 (quantile 0,99), tandis que le seuil de la méthode GEV connaît une croissance similaire mais plus importante, passant de 0,06 à 0,14 sur la même plage de quantiles. Cette progression continue montre que plus le quantile est élevé (donc plus la sensibilité de détection est accrue), plus les seuils augmentent, indiquant une exigence plus stricte pour qualifier un événement d'anomalie.

Pour un quantile donné, on constate que le seuil POT-GPD est généralement inférieur à celui de la GEV, ce qui suggère que la méthode POT-GPD est plus étroite dans la déclaration des anomalies, nécessitant des écarts moins importants par rapport à la normale.

En analysant les histogrammes, on constate que lorsque le quantile augmente (donc le seuil devient plus élevé), le nombre d'attaques détectées diminue. Ainsi, pour un quantile de 0,90, les deux méthodes détectent des centaines d'attaques. En revanche, pour un quantile de 0,99, ce nombre chute à environ de zéro attaques pour GEV et pour le POT-GPD environ de quelques dizaines. Ces résultats sont récapitulé dans le tableau III.1

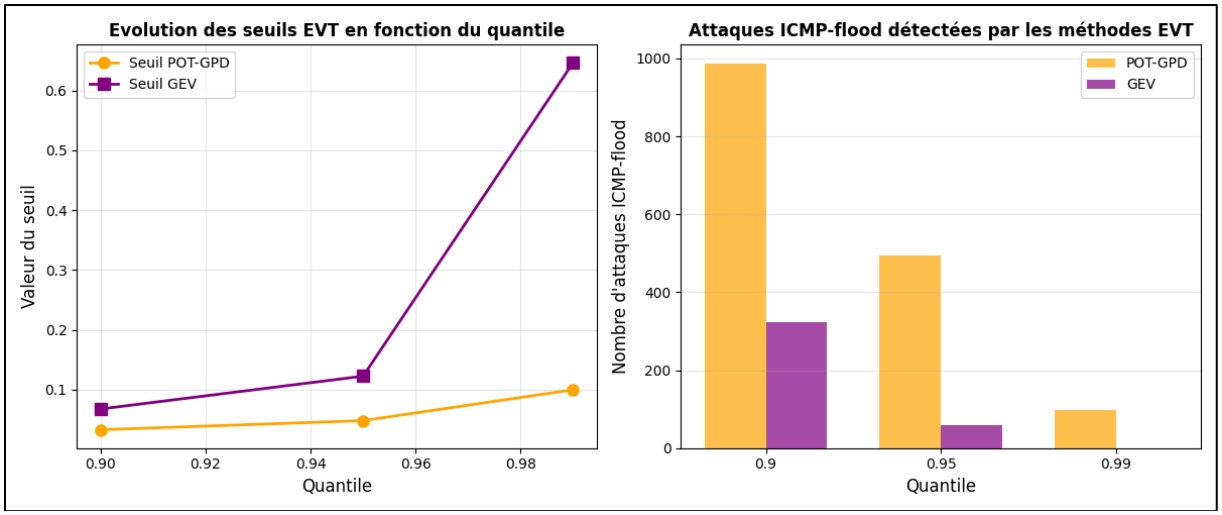

Figure III.5 : Comparaison des méthodes EVT

Quantile		0.9	0.95	0.99
POT	Seuil	0.033242	0.048478	0.100096
	Attaques détectées	987	494	99
GEV	Seuil	0.067893	0.122818	0.646703
	Attaques détectées	323	59	0

Tableau III.1. Comparaison des méthodes GEV et POT-GPD (attaques ICMP flood)

III.5.2.1.Détection les attaques UDP FLOOD

Les résultats de détection des attaques UDP Flood en environnement IPv6, utilisant les méthodes EVT POT-GPD et GEV, sont présentés respectivement dans les figures III.6, III.7 et III.8.

L'analyse des trois configurations montre que le choix du paramètre quantile détermine directement la sensibilité et l'efficacité de la méthode POT-GPD face aux attaques UDP-flood, tandis que la méthode GEV reste constamment ajustée et limitée aux événements les plus extrêmes. Avec un seuil bas Quantile = 0.90, le système se montre particulièrement robuste et détecte un grand nombre d'anomalies dès les premières augmentations de trafic (notamment entre les échantillons 0-2000 et 4200-4500), mais au coût d'un risque élevé de faux positifs. A l'opposé, un seuil très élevé à Quantile = 0.99 rend la détection extrêmement stricte, ne capturant que les pics critiques maximaux et augmentant ainsi dangereusement le risque de faux négatifs, d'autant plus que le seuil de la méthode GEV devient alors presque prohibitif. Finalement, la configuration à Quantile = 0.95 s'impose comme le compromis idéal : elle filtre les faibles fluctuations tout en ciblant efficacement les véritables pics

anormaux (notamment entre les échantillons 4200 et 5200), offrant ainsi le meilleur équilibre entre détection précoce, stabilité du réseau et réduction des fausses alertes.

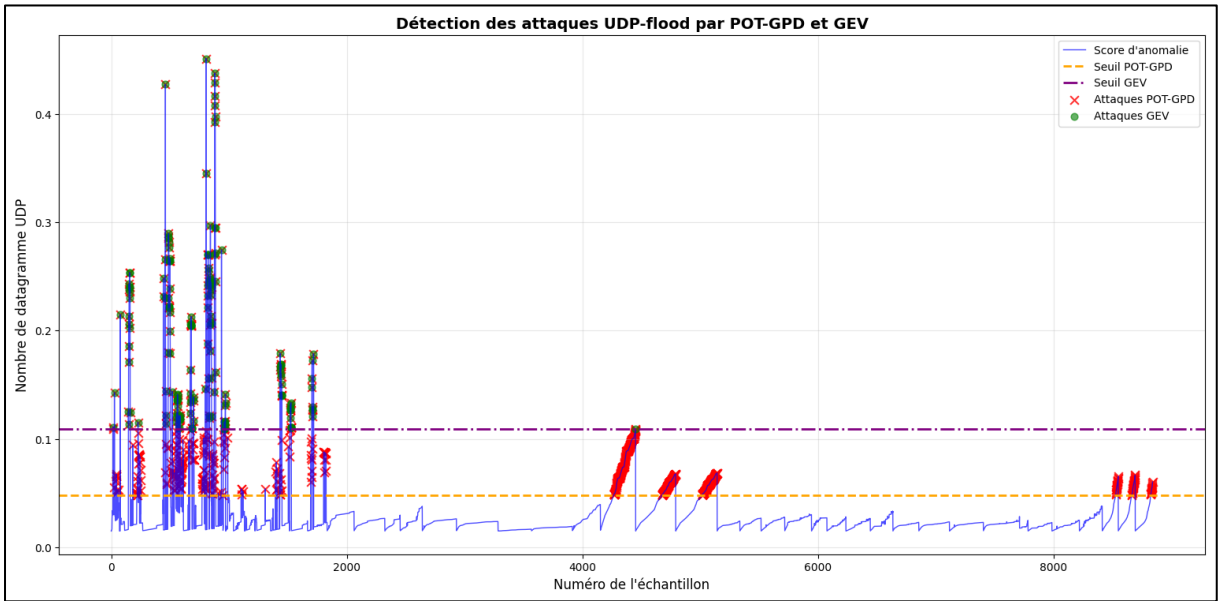

Figure III.6 : Détection des attaques UDP flood par les méthodes POT-GPD et GVE (Quantile= 0,9)

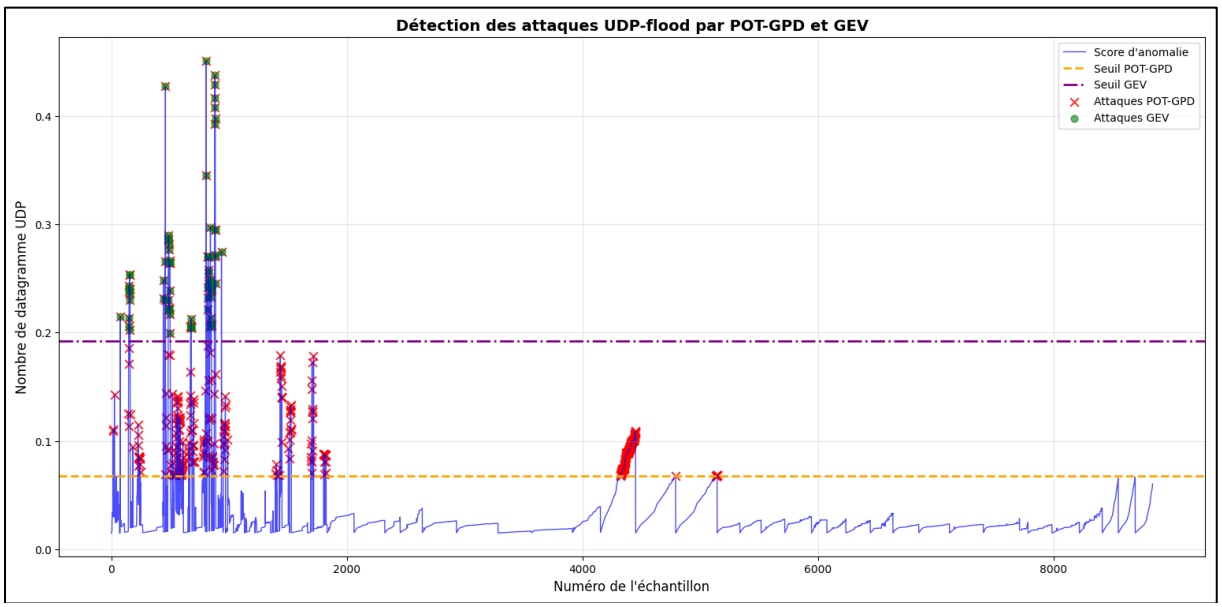

Figure III.7 : Détection des attaques UDP flood par les méthodes POT-GPD et GVE (Quantile = 0,95)

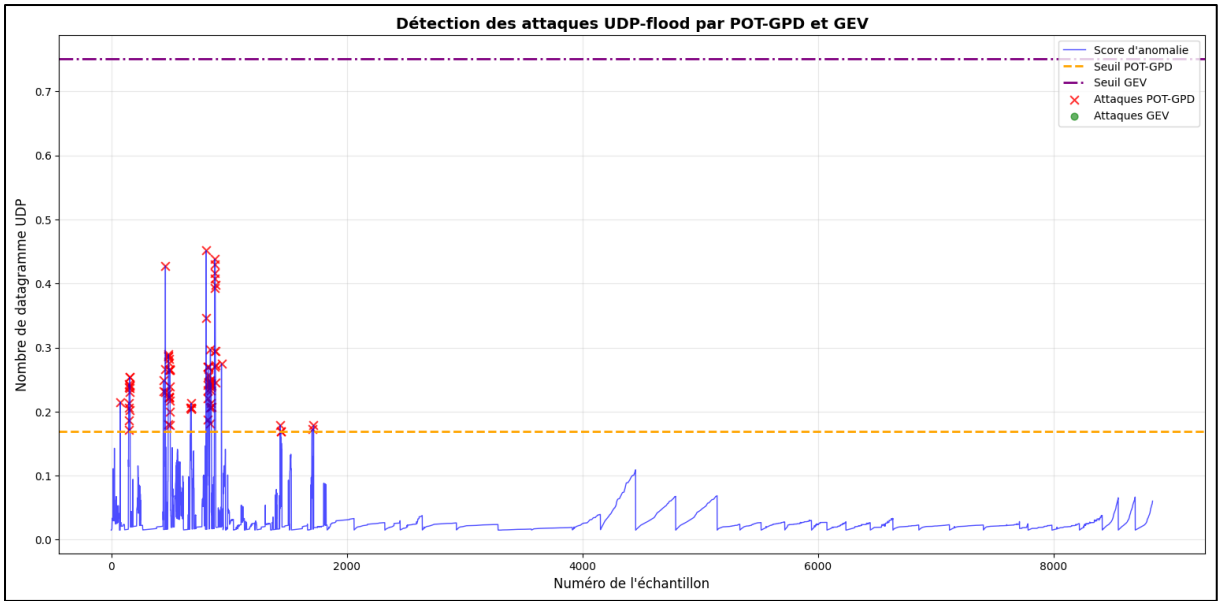

Figure III.8 : Détection des attaques UDP flood par les méthodes POT-GPD et GVE (Quantile = 0,99)

III.5.2.2.Comparaison de l'évolution des seuils EVT et du nombre d'attaques

La figure III.9et le tableau III.2 reportent les résultats de comparaison des deux méthodes sous conditions de l'UDP flood. D'une part, on observe que les seuils augmentent progressivement lorsque le quantile passe de 0.90 à 0.99, tandis que la méthode GEV produit des seuils beaucoup plus élevés que la méthode POT-GPD pour les mêmes valeurs de quantile. D'autre part, ces résultats montrent que la méthode POT-GPD détecte systématiquement un nombre plus important d'attaques UDP-flood grâce à sa plus grande sensibilité. En revanche, la méthode GEV reste plus stricte et ne détecte principalement que les événements les plus extrêmes. Enfin, l'augmentation du quantile entraîne une diminution importante du nombre d'attaques détectées pour les deux méthodes, car les seuils deviennent de plus en plus restrictifs.

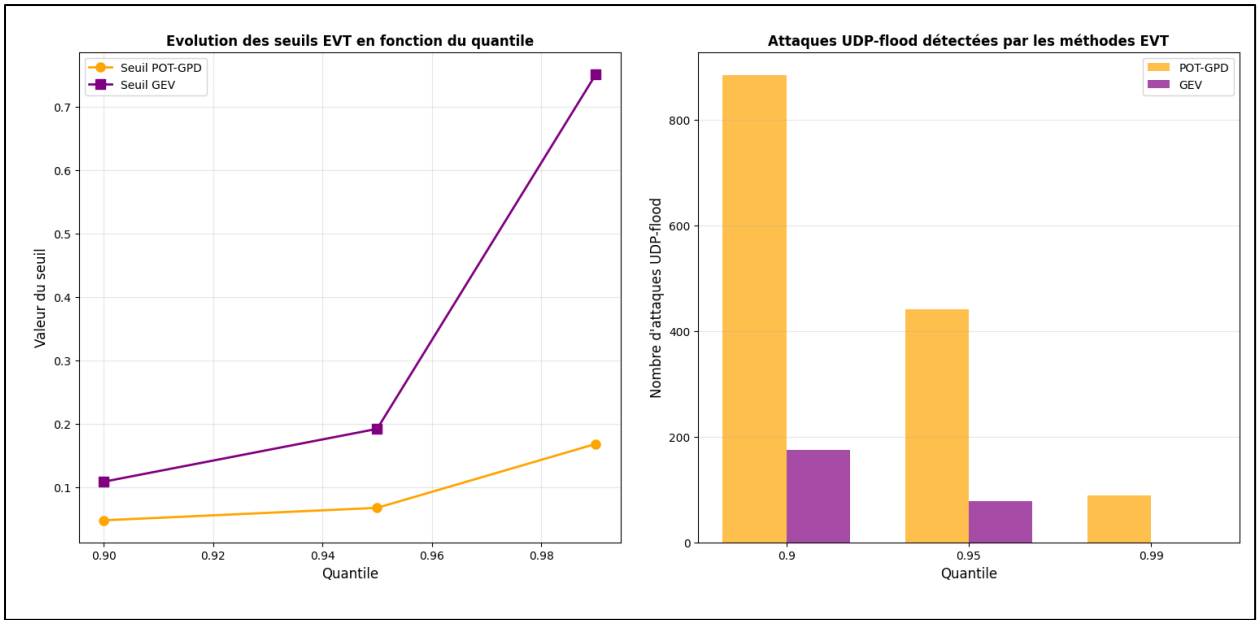

Figure III.9 : Comparaison des méthodes EVT (attaques UDP flood)

Quantile		0.9	0.95	0.99
POT	Seuil	0.048101	0.067774	0.168565
	Attaques détectées	884	442	89
GEV	Seuil	0.108950	0.192276	0.751179
	Attaques détectées	176	78	0

Tableau III.2. Comparaison des méthodes GEV et POT-GPD (attaques UDP flood)

III.5.3.1.Détection les attaques SYNFLOOD

Les figures III.10, III.11 et III.12 illustrent le comportement des approches POT-GPD et GEV face aux attaques SYN Flood à mesure que le seuil de détection augmente. La première configuration (Quantile = 0.9) montre un modèle POT-GPD très sensible capturant de nombreuses anomalies face à une GEV d'emblée plus sélective. La deuxième (Quantile = 0.95) figure atteint un compromis optimal entre détection et réduction des faux positifs. Enfin, la troisième configuration (Quantile = 0.99) met en évidence le caractère ultra-conservateur de la GEV qui, face à un quantile élevé, devient insensible aux attaques, tandis que la méthode POT-GPD maintient sa robustesse en interceptant les pics de trafic SYN les plus critiques.

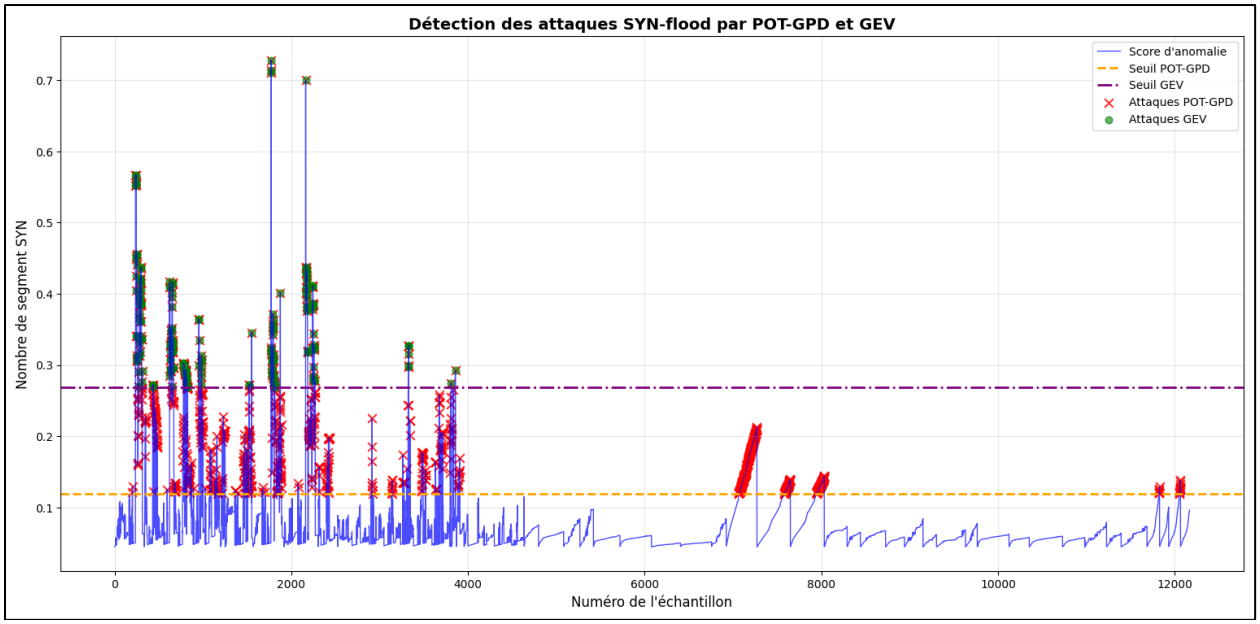

Figure III.10 : Détection des attaques SYN FLOOD par les méthodes POT-GPD et GVE (Quantile = 0,9)

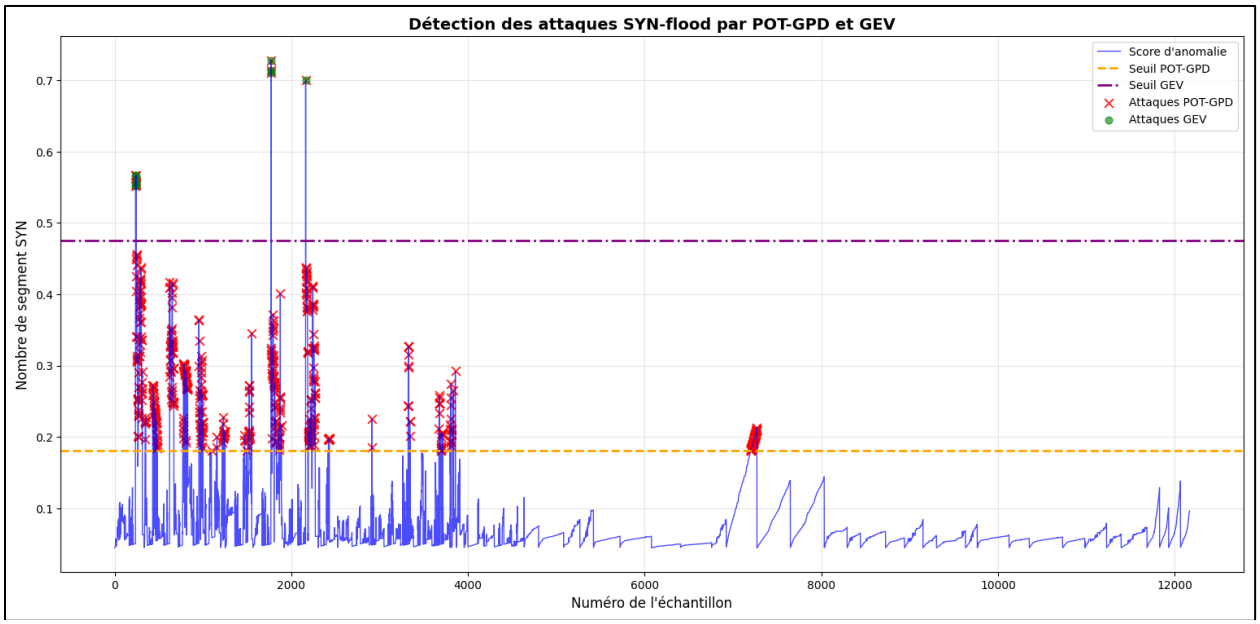

Figure III.11: Détection des attaques SYN FLOOD par les méthodes POT-GPD et GVE (Quantile= 0,95)

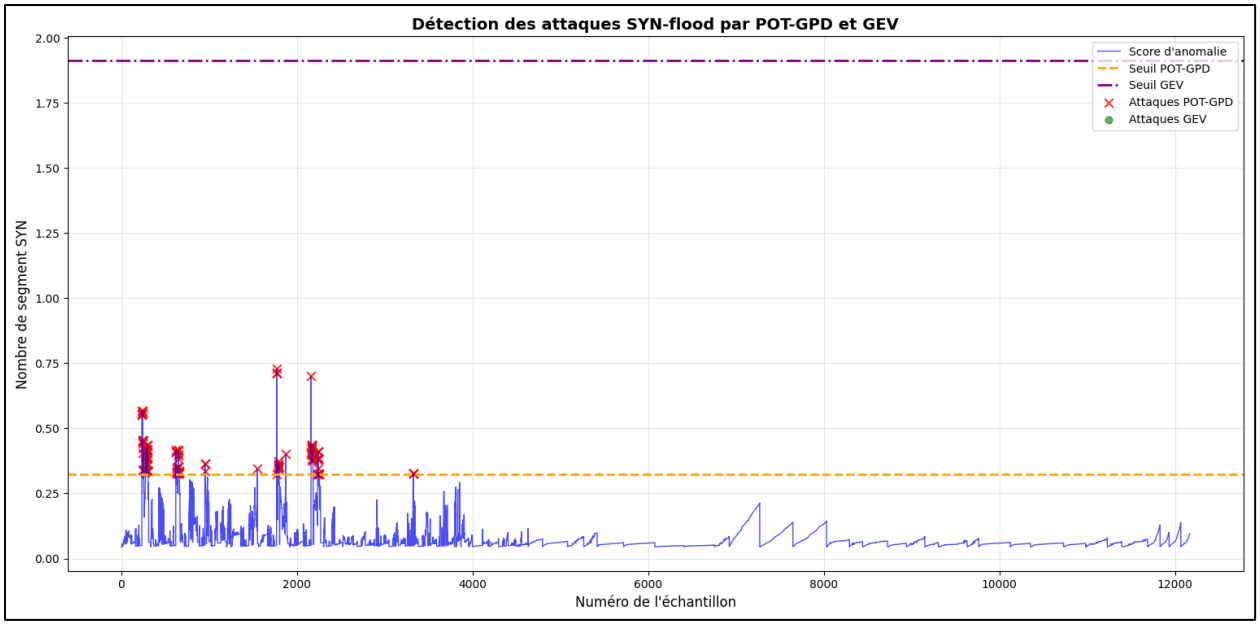

Figure III.12:Détection des attaques SYN FLOOD par les méthodes POT-GPD et GVE (Quantile = 0,99)

III.5.3.2.Comparaison de l'évolution des seuils EVT et du nombre d'attaques

La figure III.13 et le tableau III.3 présentent les résultats comparatifs des deux méthodes EVT face aux attaques SYN flood. En ce qui concerne l'évolution des seuils, on constate une augmentation progressive lorsque le quantile varie de 0,90 à 0,99. La méthode GVE se distingue par des seuils nettement plus élevés que ceux de la méthode POT-GPD pour les mêmes valeurs de quantile. Concernant la capacité de détection, la méthode POT-GPD s'avère systématiquement plus performante en détectant un plus grand nombre d'attaques SYN flood, grâce à sa sensibilité accrue. À l'inverse, la méthode GEV, plus restrictive, ne capture principalement que les événements les plus extrêmes. Enfin, l'élévation du quantile entraîne une diminution notable du nombre d'attaques détectées pour les deux approches, les seuils devenant de plus en plus exigeants.

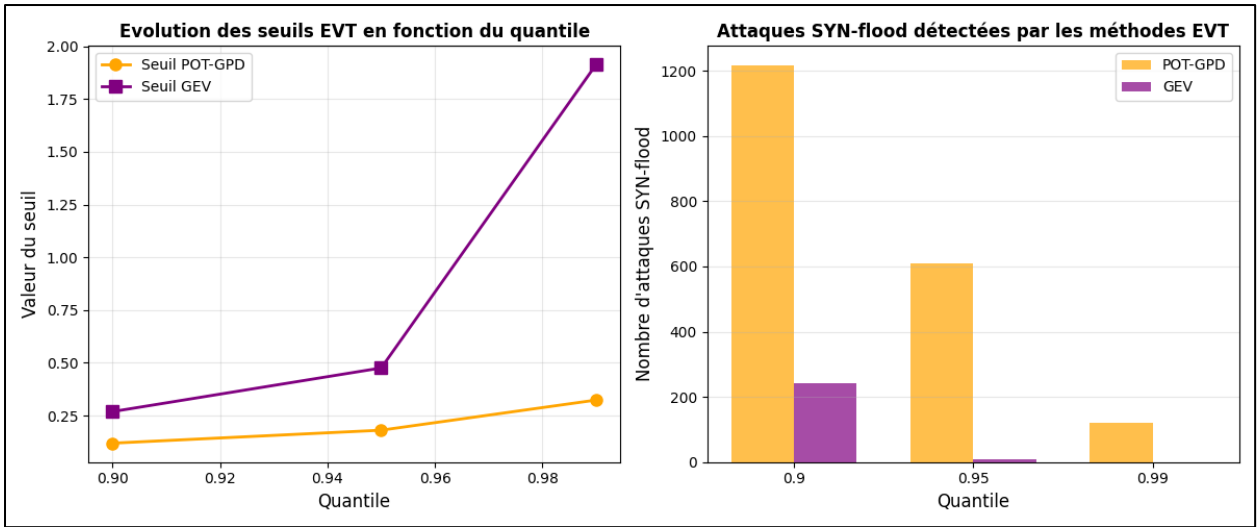

Figure III.13 : Comparaison des méthodes EVT (attaques SYN FLOOD)

Quantile		0.9	0.95	0.99
POT	Seuil	0.118908	0.180405	0.323390
	Attaques détectées	1217	609	122
GEV	Seuil	0.268560	0.475451	1.914658
	Attaques détectées	241	10	0

Tableau III.3. Comparaison des méthodes GEV et POT-GPD (attaques SYN flood)

III.5.4.1.Détection les attaques SDNDDOS

Ici, nous évaluons l'efficacité des méthodes EVT sous conditions de attaques SDN DDoS. L'analyse des figures III.14, III.15 et III.16 met en évidence l'impact du choix des quantiles sur le comportement des modèles : le quantile 0.90 engendre une hyper-sensibilité saturée de faux positifs, tandis que le quantile 0.95 offre un compromis optimal en nettoyant le bruit de fond tout en préservant la précision des détections. Enfin, le seuil extrême de 0.99 marque un point de rupture mathématique où la rigidité de la GEV provoque une non-détection totale, alors que le modèle POT-GPD confirme sa robustesse en maintenant un calibrage stable, capable d'intercepter efficacement les pics critiques de l'attaque SDNDDOS.

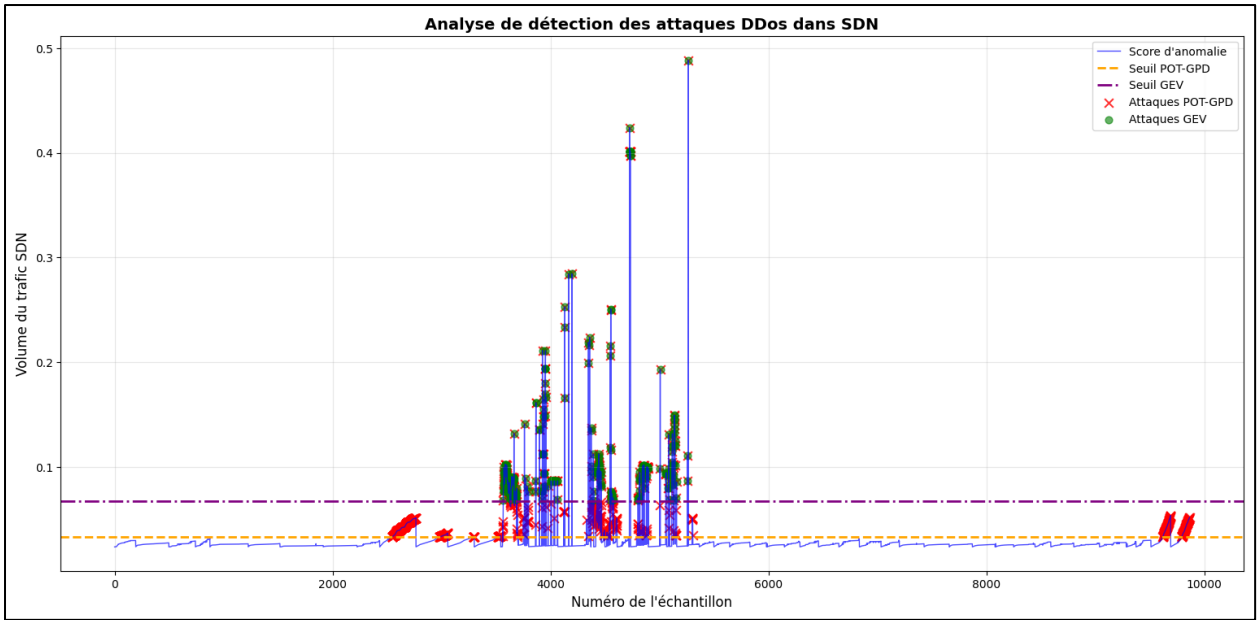

Figure III.14: Détection des attaques SND DDOS par les méthodes POT-GPD et GVE (Quantile= 0,9)

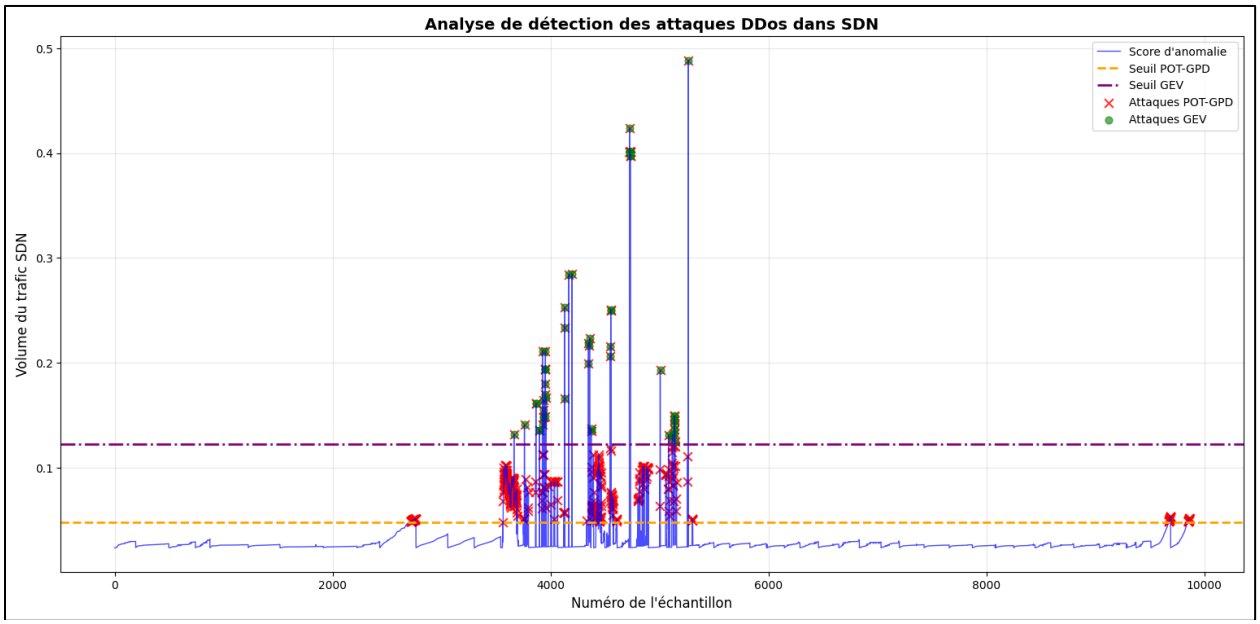

Figure III.15 :Détection des attaques SND DDOSpar les méthodes POT-GPD et GVE (Quantile = 0,95)

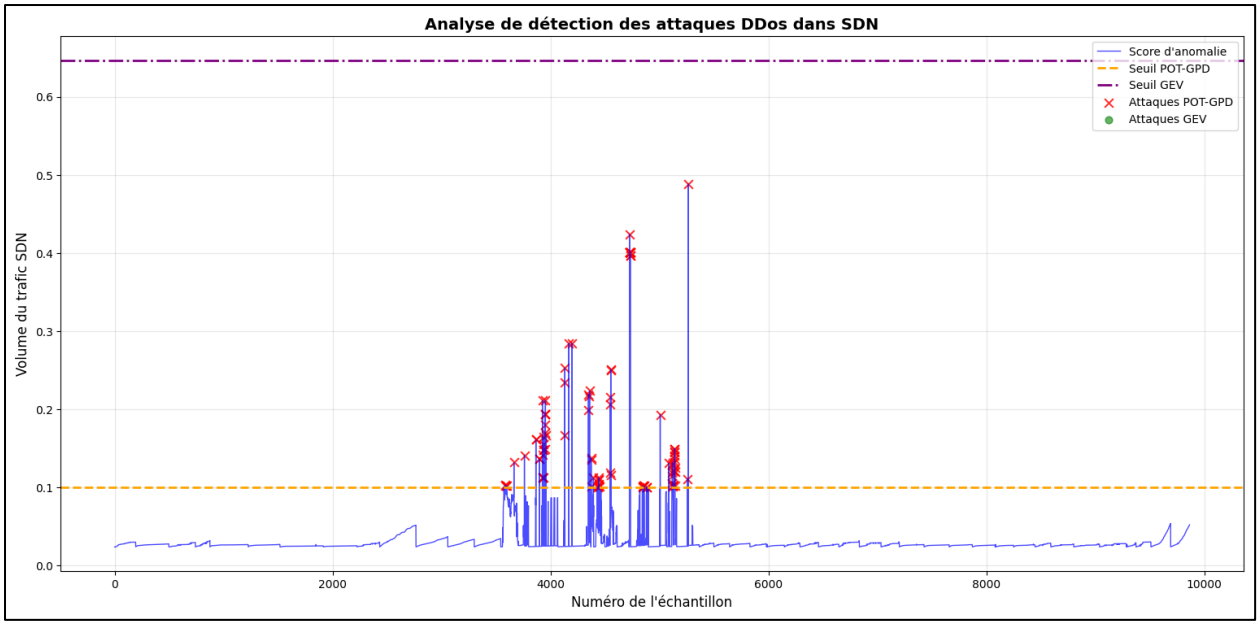

Figure III.16 : Détection des attaques SND DDOS par les méthodes POT-GPD et GVE (Quantile = 0,99)

III.5.4.2.Comparaison de l'évolution des seuils EVT et du nombre d'attaques

La figure III.17 et le tableau III.4 présentent les résultats comparatifs des deux méthodes EVT face aux attaques SDN DDOS. Leur analyse illustre qu'avec le passage du quantile de 0.90 à 0.99, le nombre d'attaques détectées par POT-GPD décroît de manière progressive (passant de près de 1000 à environ 100), tandis que les détections de GEV s'effondrent brutalement pour atteindre zéro à 0.99.

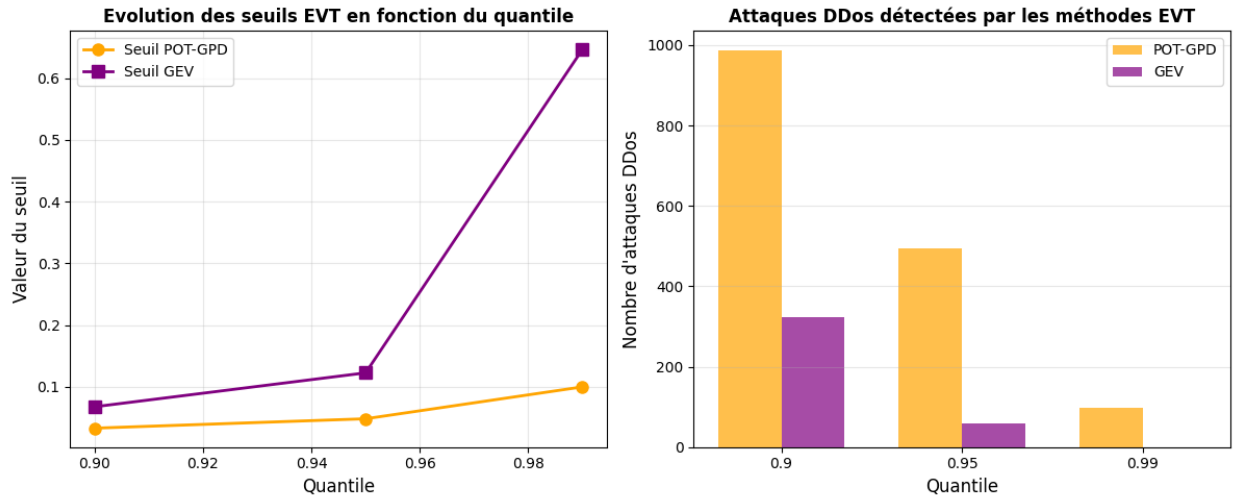

Figure III.17 : Comparaison des méthodes EVT (attaques SDN DDOS)

Quantile		0.9	0.95	0.99
POT	Seuil	0.033242	0.048478	0.100096
	Attaques détectées	987	494	99
GEV	Seuil	0.067893	0.122818	0.646703
	Attaques détectées	323	59	0

Tableau III.4. Comparaison des méthodes GEV et POT-GPD (attaques SDN DDoS)

3.5.5.1.Détection les attaques NAS

Les trois figures III.18, III.19 et III,20 illustrent la sensibilité de la détection d'anomalies selon le choix des seuils. L'approche POT-GPD reste excessivement basse sur l'ensemble des figures, ce qui lui fait systématiquement confondre le bruit de fond avec des attaques. A l'inverse, la méthode GEV s'avère bien plus précise mais dépend fortement du quantile choisi : idéale dans les deux configurations (0.9 et 0.95) pour isoler parfaitement les 16 pics critiques, elle devient totalement inefficace dans la troisième configuration (0.99) car son seuil dépasse le score maximal, rendant le modèle aveugle à toute anomalie.

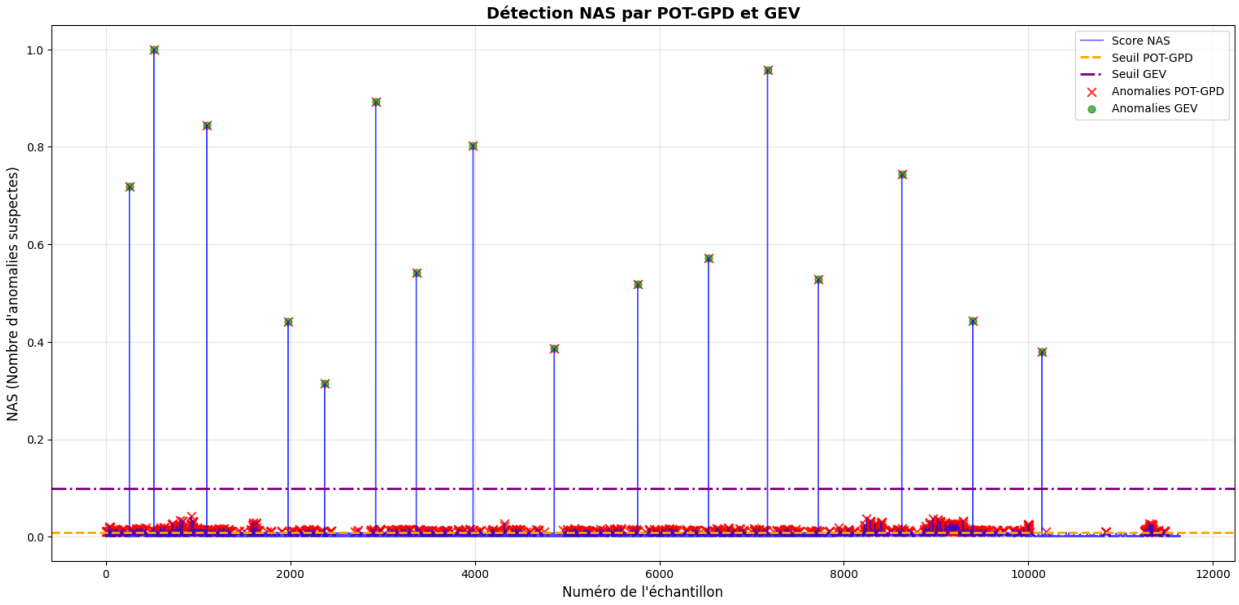

Figure III.18 : Détection des attaques NAS par les méthodes POT-GPD et GVE (Quantile= 0,9)

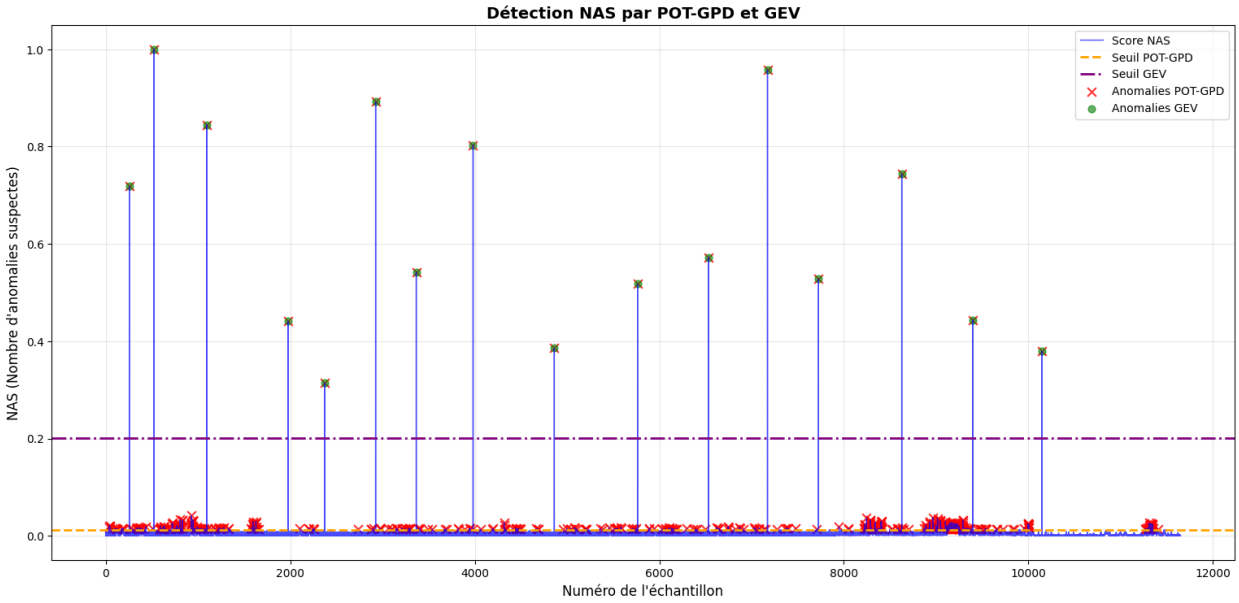

Figure III.19 : Détection des attaques NAS par les méthodes POT-GPD et GVE (Quantile = 0,95)

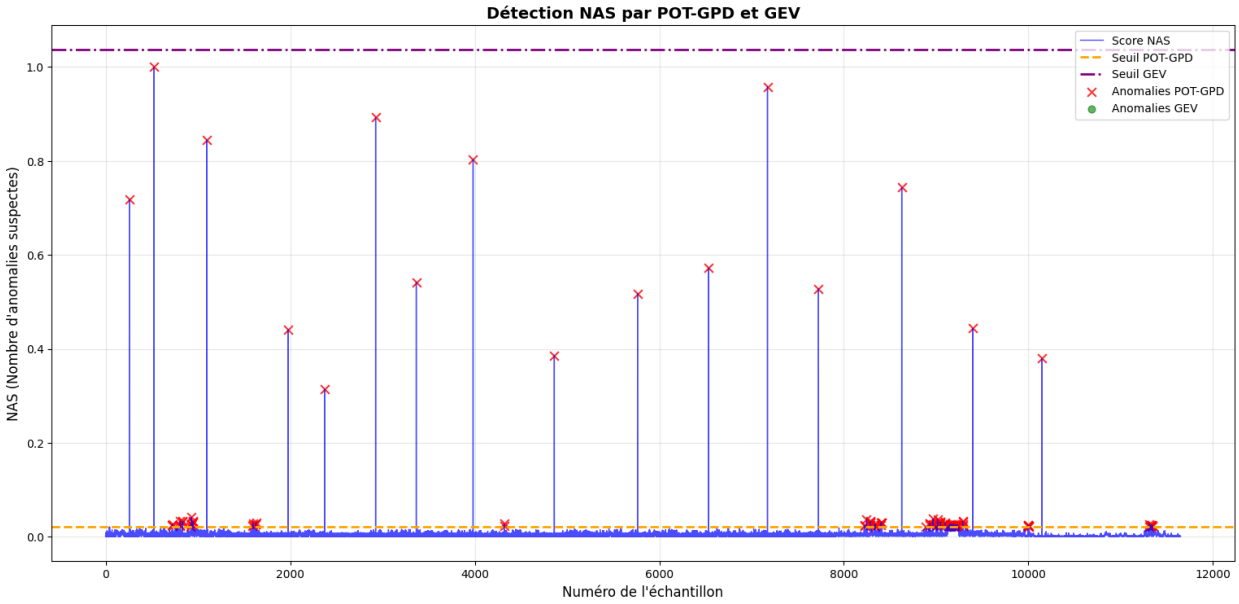

Figure III.20 : Détection des attaques NAS par les méthodes POT-GPD et GVE (Quantile = 0,99)

3.5.5.2.Comparaison de l'évolution des seuils EVT et du nombre d'attaques

La figure III.21 illustre la manière dont les valeurs du seuil et du nombre des attaques NAS détectées évoluent en fonction des quantiles choisis (0,90, 0,95, 0,99). On observe que la méthode GEV (courbe violette) impose une croissance très forte et exponentielle du seuil à mesure que le quantile augmente, passant d'une valeur proche de 0,1 à plus de 1,0. A l'inverse, la méthode POT-GPD (courbe orange) maintient un seuil beaucoup plus bas et quasi stable, augmentant très légèrement, ce qui suggère une approche beaucoup plus adaptée et sensible pour définir ce qui constitue une valeur extrême.Le nombre d'anomalies suspectes (NAS) détectées par chaque méthode selon le quantilerévèle une différence majeure d'efficacité : la méthode POT-GPD (barres orange) détecte un nombre significativement plus élevé d'anomalies que la méthode GEV (barres violettes). Pour le quantile 0,90, par exemple, POT-GPD identifie plus de 1100 anomalies, alors que GEV en identifie une quantité négligeable. On constate également que, pour les deux méthodes, le nombre d'anomalies détectées diminue à mesure que le quantile augmente, mais l'écart de performance entre les deux approches reste très marqué.

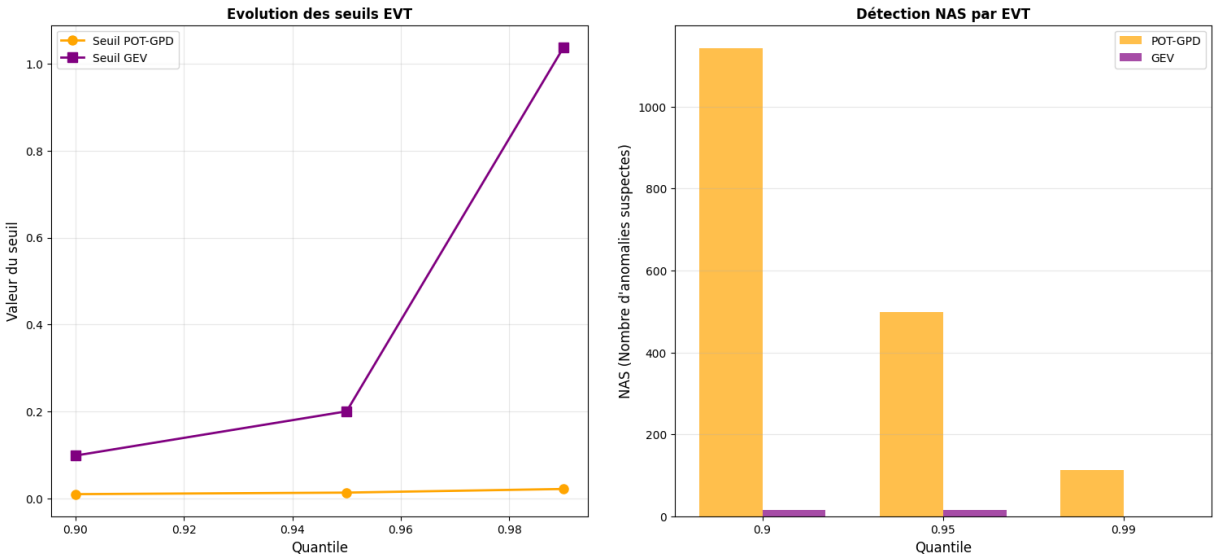

Figure III.21 : Comparaison des méthodes EVT (attaques NAS)

Quantile		0.9	0.95	0.99
POT	Seuil	0.009518	0.013087	0.021416
	Attaques détectées	1142	499	114
GEV	Seuil	0.098317	0.200175	1.037571
	Attaques détectées	16	16	0

Tableau III.5. Comparaison des méthodes GEV et POT-GPD (attaques NAS)

III.5.6.1.Détection les attaques NAD

Les figures III.22, III.23 et III.24 illustrent l'application des méthodes EVT pour détecter des anomalies dans le nombre messages NAD (ie ; attaques NAD en comparant les deux approches GEV et POT-GPD. La différence fondamentale entre ces méthodes réside dans la sévérité du filtrage : alors que le seuil POT-GPD (très bas) est extrêmement sensible et identifie une large gamme de pics comme des anomalies (indiquées par des croix rouges), le seuil GEV est beaucoup plus grand (variable selon le choix du bloc sur les trois figures). En ajustant le niveau de ce seuil, l'approche GEV permet de réduire drastiquement les faux positifs en ne ciblant que les événements les plus extrêmes (points verts), offrant ainsi une flexibilité précieuse aux administrateurs réseau pour calibrer leur système de détection selon qu'ils privilégient la sensibilité (POT) ou la spécificité (GEV) face aux menaces IPv6.

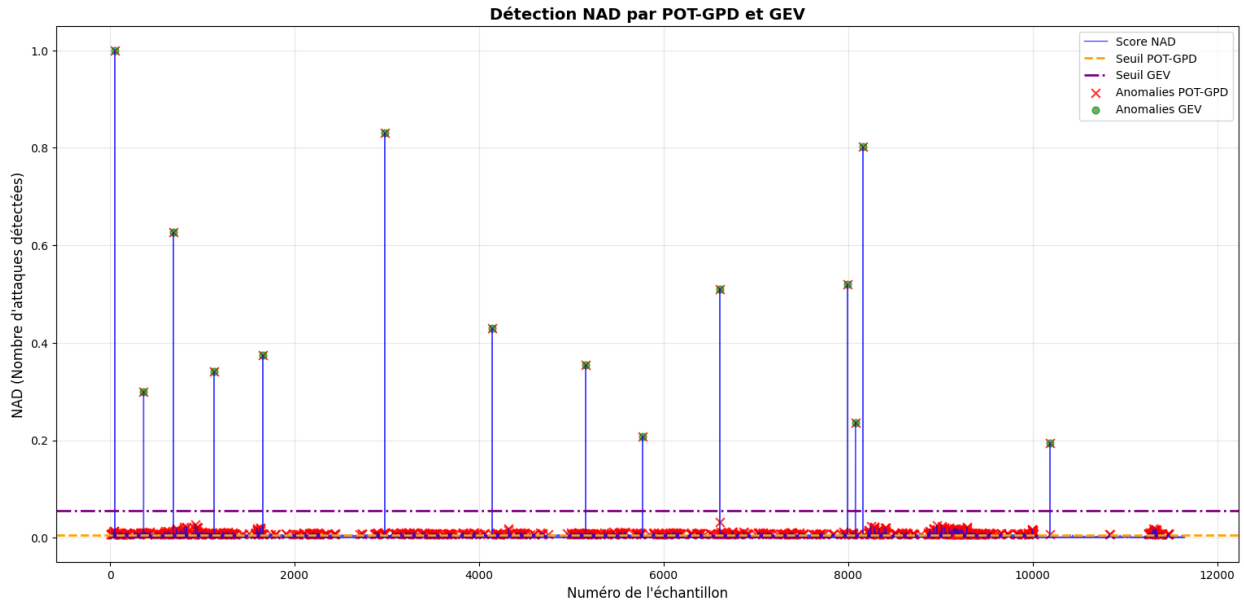

Figure III.22 :Détection des attaques NAD par les méthodes POT-GPD et GVE (Quantile= 0,9)

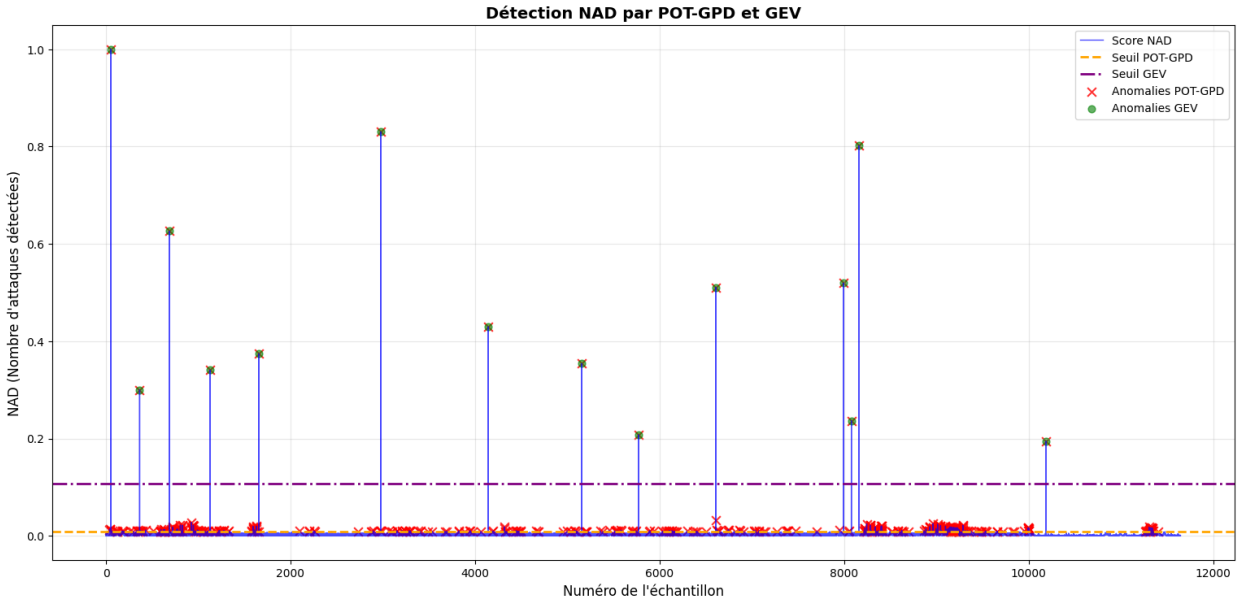

Figure III.23 :Détection des attaques NAD par les méthodes POT-GPD et GVE (Quantile = 0,95)

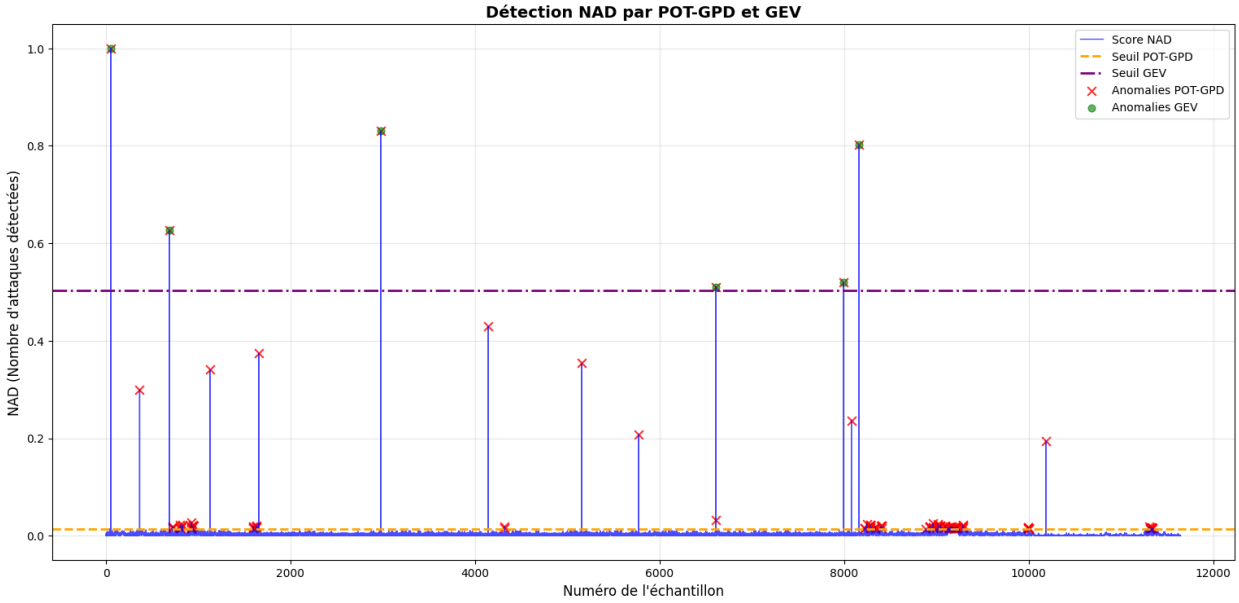

Figure III.24 :Détection des attaques NAD par les méthodes POT-GPD et GVE (Quantile = 0,99)

III.5.6.2.Comparaison de l'évolution des seuils EVT et du nombre d'attaques

La figure III.25 et le tableau III.5 comparent les deux méthode EVT en présence des attaques NAD. La courbe POT-GPD montre une augmentation faible et progressive des seuils lorsque le quantile augmente. En revanche, la courbe GEV présente une croissance beaucoup plus importante, particulièrement au quantile 0,99 où le seuil devient très élevé. Cette différence indique que la méthode GEV applique des seuils plus stricts pour identifier les événements extrêmes, tandis que POT-GPD conserve des seuils plus faibles et plus stables.

Les barres associées à POT-GPD montrent un nombre très élevé d'attaques détectées, bien que ce nombre diminue progressivement avec l'augmentation du quantile. A l'inverse, la méthode GEV détecte un nombre beaucoup plus faible d'attaques pour tous les quantiles étudiés. Cela montre que POT-GPD possède une meilleure sensibilité de détection des anomalies réseau, alors que GEV est plus restrictive à cause de ses seuils élevés.

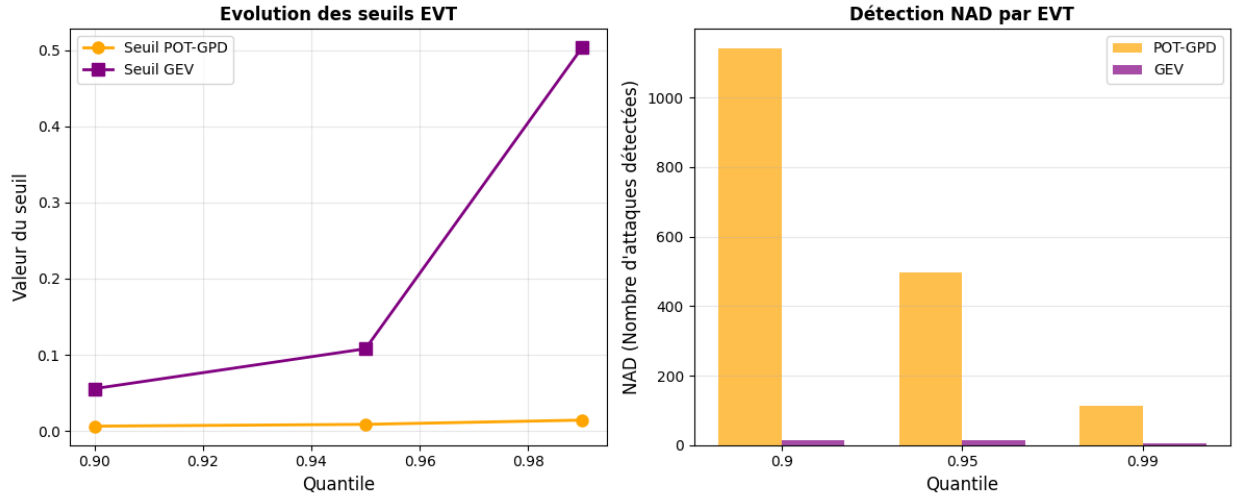

Figure III.25 : Comparaison des méthodes EVT (attaques NAD)

Quantile		0.9	0.95	0.99
POT	Seuil	0.006369	0.008758	0.014331
	Attaques détectées	1142	498	113
GEV	Seuil	0.055614	0.107953	0.503467
	Attaques détectées	14	14	6

Tableau III.6. Comparaison des méthodes GEV et POT-GPD (attaques NAD)

III.5.7.1.Détection les attaques RAD

Dans cette dernière étape, nous examinons les performances des deux méthode EVT pour la détection des attaques RAD. Les résultats obtenus sont montés dans sur les figures III.26, III.27 et III.28. La méthode POT-GPD présente un seuil faible et très sensible, permettant de détecter un grand nombre d'attaques, y compris les faibles variations du trafic. Elle offre ainsi une détection proactive mais peut générer davantage de faux positifs. A l'inverse, la méthode GEV applique des seuils plus élevés et plus sélectifs, ne conservant que les événements les plus extrêmes et critiques.

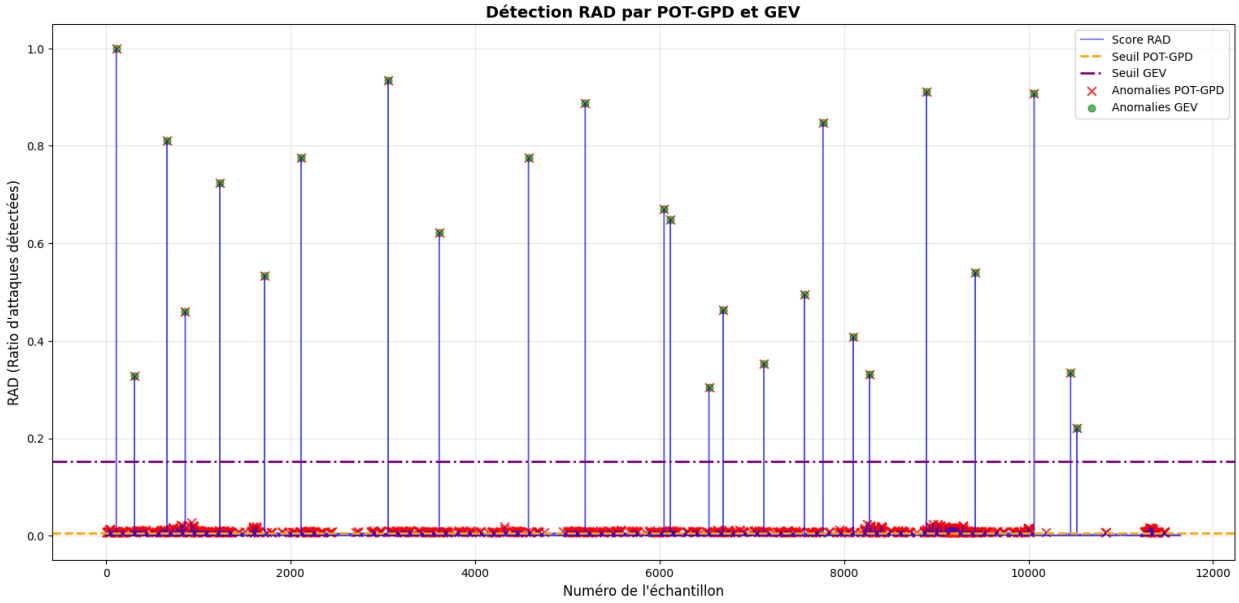

Figure III.26 :Détection des attaques RAD par les méthodes POT-GPD et GVE (Quantile= 0,9)

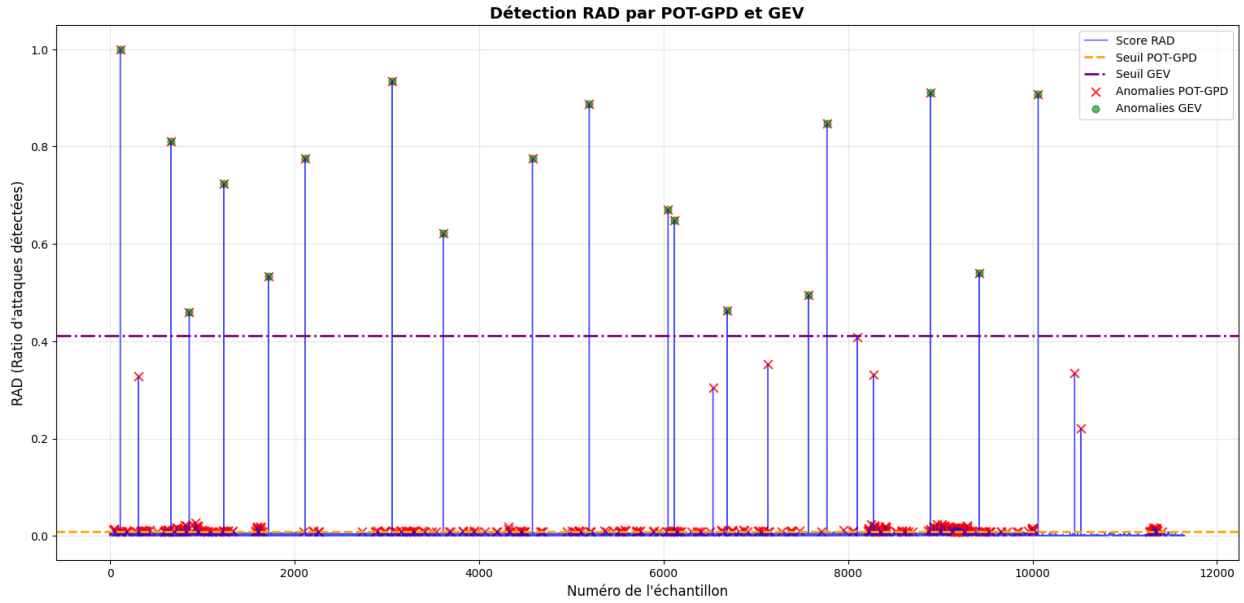

Figure III.27 :Détection des attaques RAD par les méthodes POT-GPD et GVE (Quantile = 0,95)

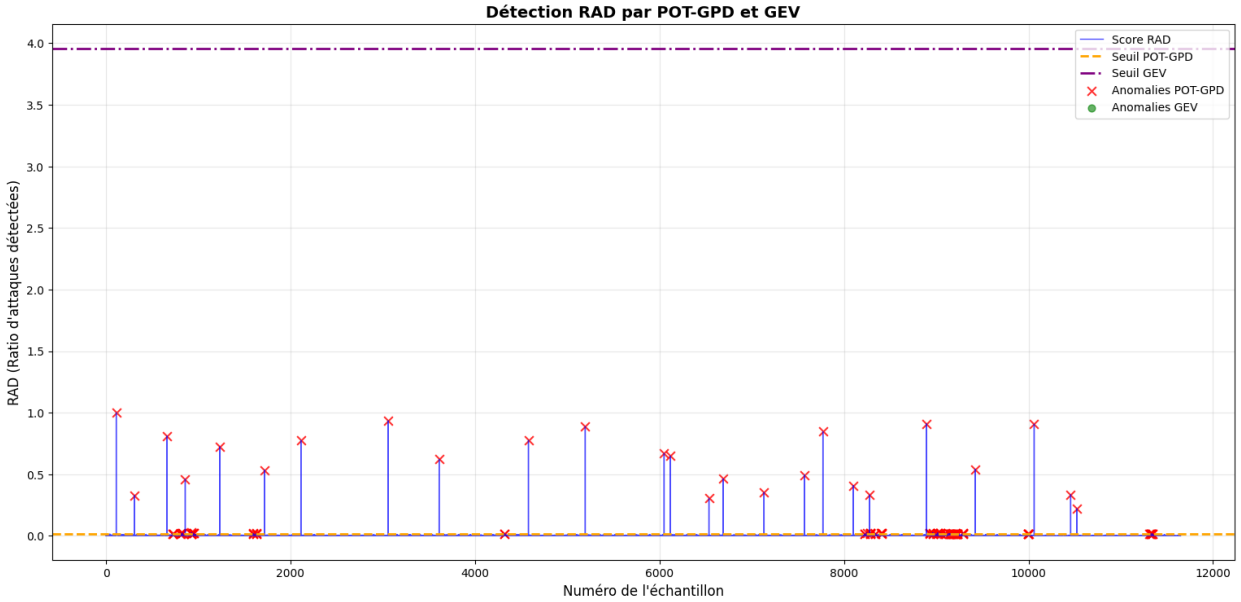

Figure III.28 : Détection des attaques RAD par les méthodes POT-GPD et GVE (Quantile =0,99)

III.5.7.2.Comparaison de l'évolution des seuils EVT et du nombre d'attaques

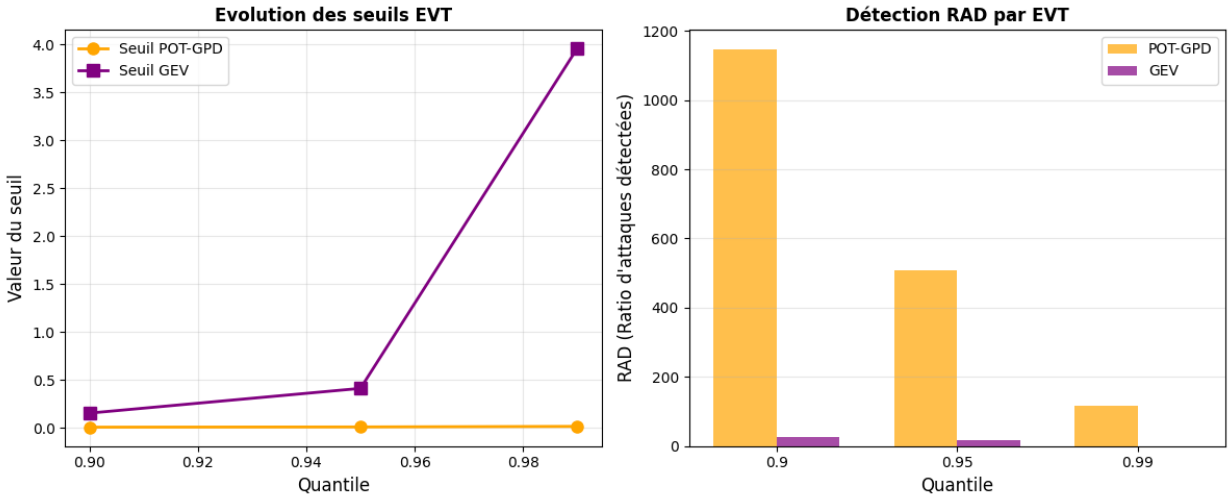

Figure III.29 : Comparaison des méthodes EVT (attaques RAD)

La figure III.29 et le tableau III.6 présentent l'analyse comparative des deux méthodes. On observe une montée en flèche du seuil GEV (violet) à mesure que le quantile augmente, témoignant de sa sélectivité croissante. A l'inverse, le seuil POT-GPD (orange) reste plat et proche de zéro, indiquant une stabilité de sensibilité.

Le diagramme en barres montre que l'approche POT-GPD détecte systématiquement un nombre bien plus élevé d'anomalies que le GEV, quel que soit le quantile. Cette figure confirme que le POT est idéal pour une surveillance exhaustive, tandis que le GEV agit comme un filtre drastique pour isoler uniquement les menaces les plus significatives.

Quantile		0.9	0.95	0.99
POT	Seuil	0.006175	0.008491	0.014481
	Attaques détectées	1147	507	117
GEV	Seuil	0.152837	0.411041	3.959140
	Attaques détectées	25	18	0

Tableau III.7. Comparaison des méthodes GEV et POT-GPD (RAD)

III.6.Conclusion

Les résultats comparatifs des méthodes POT-GPD et GEV montrent une nette supériorité de la première approche. Le POT-GPD se caractérise généralement par un seuil plus bas que celui de la GEV, ce qui lui confère une meilleure capacité à détecter un plus grand nombre d'attaques tout en maintenant un taux de fausses alarmes raisonnable. L'analyse révèle également des différences de sensibilité : le POT-GPD réagit plus efficacement aux variations locales du trafic, alors que la GEV propose une détection plus globale des anomalies. Ce chapitre confirme ainsi la pertinence de l'EVT pour la détection des cyberattaques en environnement IPv6 et souligne l'importance cruciale du réglage des paramètres pour optimiser les performances de détection.

Conclusion Générale

Conclusion Générale

Dans ce travail, nous avons étudié l'application de la théorie des valeurs extrêmes (EVT) à la détection des cyberattaques dans les réseaux IPv6. Face à l'augmentation des menaces informatiques et à la complexité croissante des flux réseau, les méthodes classiques de détection montrent souvent leurs limites, notamment lorsqu'il s'agit d'identifier des événements rares mais critiques tels que les attaques de type DDoS, ICMP Flood ou encore les intrusions massives.

L'approche EVT, basée principalement sur les modèles POT-GPD et GEV, s'est révélée particulièrement pertinente pour modéliser les queues de distribution des données réseau et identifier les comportements anormaux. Les résultats expérimentaux obtenus montrent que le choix du seuil et du quantile influence directement le nombre d'attaques détectées, ce qui confirme l'importance d'un bon paramétrage pour équilibrer sensibilité et précision de la détection.

La comparaison entre les deux approches met en évidence que la méthode POT-GPD est généralement plus flexible et plus sensible aux variations locales des données, tandis que la méthode GEV offre une vision globale des extrêmes sur l'ensemble de l'échantillon. Cette complémentarité permet d'améliorer la robustesse du système de détection.

Ainsi, l'intégration de la théorie des valeurs extrêmes dans les systèmes de cybersécurité constitue une solution efficace pour renforcer la détection précoce des anomalies dans les réseaux IPv6. Toutefois, cette approche reste dépendante de la qualité des données et du choix des paramètres statistiques, ce qui ouvre la voie à de futures améliorations, notamment par l'intégration de techniques d'apprentissage automatique ou d'adaptation dynamique des seuils.

En perspective, il serait intéressant d'étendre cette étude à des environnements réseau plus complexes et en temps réel, afin de valider la performance de ces modèles dans des conditions opérationnelles réelles.

Références

[1] https://www.coursera.org/fr-FR/articles/ipv4-vs-ipv6

[2] http://cisco.ofppt.info/ccna1/course/module6/6.1.4.4/6.1.4.4.html

[3] https://www.rfc-editor.org/info/rfc8200/#section-3

[4] https://docs.oracle.com/cd/E19957-01/820-2982/ipv6-planning-7/index.html

[5] https://www.google.fr/ipv6/statistics.html

[6] https://www.cyber.gc.ca/fr/orientation/considerations-securite-protocole-internet-version-6-itsm80003

[7] https://facntic.univ-constantine2.dz/wp-content/uploads/2020/10/Support-Cours-SERE.1.pdf

[8] https://datatracker.ietf.org/doc/html/rfc4443#page-3

[9] https://www.omnisecu.com/tcpip/ipv6/icmpv6-functions-icmpv6-packet-format-and-icmpv6-message-types.php

[10] http://livre.g6.asso.fr/index.php/ICMPv6#inaccessible

[11] https://connect.ed-diamond.com/MISC/misc-072/protections-de-couche-2-contre-les-attaques-visant-les-reseaux-ipv6

[12] https://www.akamai.com/fr/glossary/what-is-icmp-flood-ddos-attack

[13] https://biblio.univ-annaba.dz/wp-content/uploads/2022/12/These-Dehaimi-Samira.pdf

[14] Smith, R., (1985), Maximum likelihood estimation in a class of non regular cases, Biometrika, 72, 6790.

[15] Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. Water Resources Research, 15(5), 1049-1054.

[16] Christopeit, N. (1994). Method of moments for semi parametric location-scale models. Journal of Statistical Planning and Inference, 38(2), 167-185.

[17] Hosking.J.R.M, (1990). L-moments: analysis and estimation of distributions using linear combinations of statistics, Journal of the Royal Statistical Society.105 124.

[18] Pickands III, J., et al., (1975). Statistical inference using extreme order statistics. The Annals of Statistics 3 (1), 119 131.

[19] Hill, B. (1975). A Simple Approach to Inference About the Tail of a Distribution. Annals of Statistics 3, 1163-1174.

[20] Dekkers, A. L., Einmahl, J. H., De Haan, L., (1989). A moment estimator for the index of an extreme-value distribution. The Annals of Statistics, 1833 1855.

Références

[21] Benelmir.I. (2010). Les L-moments : Application en hydrologie. magister en mathématique. Univ-biskra.

[22] De Haan, L. and Ferreira, A., 2006. Extreme Values Theory: An introduction. Springer

[23] Davis, R. and Resnick, S.I. (1984). Tail Estimates Motivated by Extreme Value Theory. Annals of Statistics 12, 1467-1487.

[24] Hausler, E. and Teugels, J.L. (1985). On Asymptotic Normality of Hill s estimator for the Exponent of Regular Variation. Annals of Statistics 13, 743-756

[25] Mason, D.M. (1982). Laws of Large Numbers for Sums of Extreme Values. Annals of Probability 10, 756-764.

[26] Deheuvels, P., Hausler, E. and Mason, D.M. (1988). Almost Sure Convergence of the Hill Estimator. Mathematical Proceedings of the Cambridge Philosophical Society 104, 371-381.

[27] Drees, H.(1996). Rened Pickands Estimator of the Extreme Value Index. Annals of Statistics 23, 2059-2080.

[28] Drees, H. and Kaufmann, E. (1998). Selection of the Optimal Sample Fraction in Univariate Extreme Value Estimation. Stochastic Processes and their Applications 75, 149-195.

[29] Dekkers, A. L., De Haan, L., et al., (1989). On the estimation of the extreme value index and large quantile estimation. The Annals of Statistics 17 (4), 1795 1832.

[30] https://data.mendeley.com/datasets/x6vr3sdm75/1