Detection d'attaques DDoS dans le contexte d'un fournisseur cloud de grande envergure

Clément Boin

THÈSE

Présentée et soutenue publiquement le
18 Décembre 2023

Pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITÉ DE LILLE
spécialité Informatique

par
Clément Boin

Détection d'attaques DDoS dans le contexte d'un fournisseur cloud de grande envergure

COMPOSITION DU JURY :

Gilles GRIMAUD // *Professeur des universités - Université de Lille* // **Directeur de thèse**
David BROSSET // *Maître de conférences HDR - École Navale, IReNav, Arts et Métiers Sciences & Technologies* // **Rapporteur**
Etienne RIVIÈRE // *Professeur des universités - UCLouvain* // **Rapporteur**
Vania MARANGOZOVA // *Maîtresse de conférences - Université de Grenoble Alpes* // **Examinatrice**
Giuseppe LIPARI // *Professeur des universités - Université de Lille* // **Examinateur et Président du jury**
Michaël HAUSPIE // *Maître de conférences HDR - Université de Lille* // **Co-directeur de thèse**
Tristan GROLÉAT // *Docteur - OVHcloud* // **Invité**

CRIStAL - UMR 9189
ED MADIS - 631

Table des matières

Résumé

L'objet de cette thèse est la conception et le développement d'un système de détection des attaques DDoS volumétriques, intégré au sein d'une infrastructure en nuage (*cloud*). Cette nouvelle proposition vise à remplacer un système préexistant qui a été jugé peu adaptable et complexe à exploiter par les ingénieurs d'OVHcloud. Afin d'atteindre cet objectif, le travail de thèse est articulé autour de quatre axes majeurs.

Tout d'abord, une revue exhaustive de la littérature scientifique est entreprise pour appréhender les problématiques associées à la détection des attaques volumétriques dans le contexte spécifique des environnements en nuage. Les attaques DDoS, depuis leur avènement au début des années 2000, n'ont cessé de gagner en sophistication et en ampleur. Les environnements tels que celui d'OVHcloud font l'objet de centaines d'attaques DDoS quotidiennes, dont certaines dépassent le seuil du téraoctet de trafic. Dans une première contribution, l'examen détaillé d'une année d'attaques visant l'infrastructure d'OVHcloud révèle que peu de travaux antérieurs tiennent compte de tels niveaux de volumétrie. Cette constatation initiale met en évidence la nécessité d'adapter les solutions de pointe existantes afin qu'elles puissent être appliquées dans des environnements à haute performance.

Dans un second volet, il est démontré que les ensembles de données disponibles pour la recherche sont peu compatibles sur le plan statistique avec les conditions observées dans le cadre de cette étude. Les métriques largement utilisées dans la littérature scientifique ne parviennent pas à capturer la réalité quotidienne. Cette lacune génère des problématiques, tant du point de vue de la conception de solutions adaptées à ce contexte spécifique que de la reproductibilité des travaux de recherche. Du point de vue des hébergeurs, l'absence de jeux de données appropriés s'explique partiellement par les difficultés rencontrées par le milieu académique pour accéder aux infrastructures industrielles, majoritairement sous l'égide de grandes multinationales du secteur privé. Les considérations liées à la confidentialité des données à caractère personnel présentes dans de tels jeux de données constituent également un frein. C'est ainsi que dans un troisième apport significatif, une proposition de générateur de trafic est formulée, respectant les propriétés statistiques spécifiques à l'infrastructure cloud étudiée.

Fort de cette compréhension accrue des problématiques internes aux fournisseurs de services en nuage pour la détection des attaques DDoS, ainsi que des défis liés à la reproduction de situations réelles, incluant à la fois le trafic nominal et les attaques, un quatrième et dernier volet, sous la forme d'un brevet industriel, est consacré à la description d'une architecture de système de détection des attaques DDoS volumétriques. Cette architecture doit permettre l'intégration d'algorithmes de détection, tout en restant maintenable par les experts métier. De plus, elle doit être conçue pour résoudre les problématiques relatives à la charge réseau générée par une infrastructure accueillant des millions de clients à travers le monde.

Abstract

The objective of this thesis is the conception and development of a system for detecting volumetric DDoS attacks, integrated within a cloud infrastructure. This novel proposition aims to supplant an existing system deemed to be inadequately adaptable and operationally complex for OVHcloud engineers. To achieve this objective, the thesis is structured around four primary axes.

Firstly, a comprehensive review of the scientific literature is undertaken to apprehend the issues associated with detecting volumetric attacks within the specific context of cloud environments. Since their emergence in the early 2000s, DDoS attacks have continually increased in sophistication and magnitude. Environments such as OVHcloud are subjected to hundreds of daily DDoS attacks, with some exceeding the terabit traffic threshold. In a primary contribution, a detailed examination of a year's worth of attacks targeting the OVHcloud infrastructure reveals that few prior works take such levels of volume into account. This initial observation underscores the necessity of adapting existing state-of-the-art solutions for application in high-performance environments.

In a secondary facet, it is demonstrated that the available datasets for research lack statistical compatibility with the observed conditions within this study's framework. Widely employed metrics in scientific literature fail to capture everyday realities. This shortfall generates issues both in terms of devising context-specific solutions and in reproducing research outcomes. From the perspective of hosting providers, the absence of suitable datasets is partially attributed to the difficulties faced by the academic community in accessing industrial infrastructures, predominantly under the purview of major private-sector multinationals. Considerations linked to the confidentiality of personally identifiable information within such datasets also impede progress. Thus, in a significant tertiary contribution, a traffic generator proposal is formulated, adhering to the specific statistical properties of the studied cloud infrastructure.

Leveraging this heightened comprehension of the intrinsic challenges faced by cloud service providers in detecting DDoS attacks, as well as the obstacles posed by the replication of real-world scenarios, encompassing both normal traffic and attacks, a fourth and final facet, presented in the form of an industrial patent, is devoted to

delineating an architecture for detecting volumetric DDoS attacks. This architecture must facilitate the integration of detection algorithms while remaining maintainable by domain experts. Furthermore, it should be designed to address issues pertaining to the network load engendered by an infrastructure accommodating millions of clients across the globe.

Remerciements

Ce document concrétise plus de trois années de travail, rendues possibles grâce à l'aide précieuse de nombreuses personnes.

Je tiens à exprimer ma profonde gratitude envers Gilles Grimaud, Michaël Hauspie et Giuseppe Lipari, qui m'ont accompagné et ont éveillé mon intérêt pour l'informatique dès mes premières années d'études supérieures. C'est notamment grâce à Michaël, depuis le DUT, qu'est née ma passion pour la recherche, à travers un stage au sein de l'équipe 2XS, et au fil des années jusqu'à la rédaction de ce document. Au cours de cette période, j'ai eu le privilège de collaborer avec des personnes qui marqueront la suite de ma carrière, telles que Pierre Graux, Thomas Vantroys, David Nowak, Damien Amara, Alexandre Boé (membre honorifique de 2XS) et Olivier Lourme, mon binôme de bureau et de démarches administratives. François Serman a également joué un rôle crucial en me permettant de rencontrer les personnes chez OVHcloud qui ont facilité la réalisation de ces travaux dans les meilleures conditions possibles avant de ce diriger vers de nouveaux horizons.

Je suis reconnaissant envers David Brosset et Étienne Rivière pour avoir rapporté ma thèse et pour leur collaboration précieuse dans l'amélioration du manuscrit jusqu'à sa version actuelle.

Je tiens à exprimer ma sincère reconnaissance envers mes collègues et amis d'OVHcloud, présents et passés, pour leur aide précieuse et leur bonne humeur tout au long de cette thèse. Mes remerciements vont à Xavier Guillaume, Bastien Dhiver, Tristan Groléat, Gautier Mathon, Clément Sciascia, Maxime Monet, Charlelie Ravail, et mon collègue de bureau direct, Christophe Bacara. J'ai redécouvert Christophe après mon stage chez 2XS, alors qu'il préparait son doctorat, ainsi que toutes les personnes de l'open-space *BattleStar*, Yaniv Fdida, Louis Declerfayt et Kevin Disneur, ainsi que l'ensemble de l'équipe *Network*.

Je souhaite également exprimer ma reconnaissance envers ma famille, en particulier envers ma mère et mon père, qui m'ont toujours soutenu dans mes études sans jamais remettre en question mes choix. Mes amis méritent également un immense merci pour leur soutien inconditionnel tout au long de mes études. Je tiens à remercier Benjamin Lopez, Thomas Campistron, Sophie Kaleba, Célestine Sauvage, Axel

Thavisouk, Florian Vanhmes, Martin Lemesle, Mehdi Malamelli, Sami Halabi, ainsi que tous ceux qui, bien que non cités, sont venus ici chercher leur prénom.

Enfin, je voudrais exprimer ma reconnaissance envers ma fiancée, Ophélie, qui m'accompagne depuis plus de onze ans, pour tous les moments heureux que nous avons partagés et que nous continuerons de partager dans les prochains mois qui promettent d'apporter de merveilleux changements dans ma vie, avec de nouvelles responsabilités et une perspective plus épanouissante qui grandit chaque jour.

Introduction

Motivations

L'informatique est devenue un pilier incontournable de notre société contemporaine, une puissance qui a profondément redéfini notre mode de vie, de travail, de divertissement, et même notre perception du monde qui nous entoure. Cette transformation a eu un impact si profond qu'elle a remodelé fondamentalement nos vies, qu'elles soient personnelles ou professionnelles, touchant chaque aspect de notre existence. Elle a révolutionné la manière dont nous interagissons, partageons des informations, et collaborons à une échelle mondiale. Nous assistons actuellement à une révolution numérique, et au cœur de cette révolution se trouvent les entreprises spécialisées dans les services informatiques, en particulier les fournisseurs de services *cloud* [SP20].

Ces acteurs occupent une position centrale dans le paysage numérique contemporain. En offrant des solutions d'hébergement et de gestion de données, ils ont conquis le marché mondial, répondant ainsi aux besoins variés et complexes d'un nombre croissant d'organisations. Désormais, entreprises, gouvernements, établissements éducatifs, professionnels de la santé, et bien d'autres encore, comptent sur ces prestataires de services *cloud* pour stocker, gérer, et distribuer leurs données de manière efficace et sécurisée.

L'évolution rapide de la technologie informatique, couplée à la croissance de la complexité des besoins numériques, a élevé ces entreprises au rang d'acteurs clés dans la concrétisation des aspirations et des ambitions de la société contemporaine. Leur impact va bien au-delà de celui de simples fournisseurs de services, car ils contribuent de manière substantielle à influencer notre manière d'appréhender les défis et les opportunités d'un monde de plus en plus interconnecté et numérisé [ZJ22].

Cependant, cette dépendance croissante à l'égard des services informatiques a, malheureusement, ouvert la porte à une menace qui ne cesse de croître et qui est omniprésente : *les cyberattaques*. Ces attaques numériques ciblent une multitude d'objectifs, allant de la simple démonstration de compétences techniques à l'extorsion financière, voire à des opérations militaires sophistiquées [LC09]. La cybersécurité

est devenue un enjeu majeur de notre ère numérique. À mesure que les entreprises et les particuliers dépendent de plus en plus des technologies de l'information pour leurs activités quotidiennes, la protection des données et des systèmes informatiques est devenue cruciale. Pour garantir une sécurité informatique efficace, il est essentiel de prendre en compte divers critères et menaces potentielles. Parmi ces critères, la résilience aux attaques *Distributed Denial of Service* (DDoS) occupe une place particulièrement importante. Les critères de la cybersécurité englobent un large éventail de mesures, de politiques et de pratiques visant à protéger les informations sensibles, à prévenir les intrusions malveillantes et à maintenir la disponibilité et l'intégrité des systèmes informatiques. Ces critères incluent la confidentialité, l'authenticité, la disponibilité, l'intégrité et la non-répudiation des données. Pourtant, même si toutes ces dimensions sont essentielles, la disponibilité des systèmes et des services est souvent négligée et pourtant cruciale. Les attaques DDoS représentent l'une des menaces les plus graves pour la disponibilité des services en ligne. Lors d'une attaque DDoS, un grand nombre de dispositifs informatiques sont compromis pour submerger un serveur ou une infrastructure réseau cible de trafic, le rendant ainsi inaccessible aux utilisateurs légitimes. [BCL21].

Les attaques DDoS représentent une forme de sabotage numérique où les assaillants coordonnent délibérément une avalanche de trafic en direction d'une cible spécifique. Ce déluge de trafic sature la capacité du réseau de la cible, entraînant invariablement une interruption de service si des mesures d'atténuation ne sont pas mises en place. Les conséquences de ces attaques peuvent être dévastatrices, provoquant d'importantes perturbations, des pertes financières considérables, et une détérioration significative de la réputation des organisations prises pour cible. De plus, au fil du temps, ces attaques ont gagné en sophistication, devenant plus furtives, complexes et répandues, ce qui les rend encore plus insidieuses [Abh19].

Il est impératif de saisir que les attaques DDoS ne se limitent pas à être un simple problème technique, mais constituent une menace transcendant les frontières et les secteurs. Elles ont le potentiel de paralyser des services essentiels, que ce soit dans les domaines financiers, de la santé, des communications, voire de la sécurité nationale. Leur nature insidieuse souligne l'urgence de développer des stratégies avancées de détection et d'atténuation pour les contrer.

Ainsi, face à la montée en puissance des attaques DDoS, il devient impératif de mobiliser la recherche académique et industrielle pour élaborer des solutions innovantes, ancrées dans une base solide de connaissances, plutôt que de compter sur des réponses improvisées ou des stratégies basées sur des impressions superficielles. Les enjeux sont bien trop importants pour être traités de manière superficielle, et il incombe aux chercheurs et aux responsables de la sécurité de jouer un rôle central dans la défense des infrastructures numériques et dans la préservation de la stabilité de notre société de plus en plus interconnectée [BN20].

C'est précisément la combinaison complexe d'enjeux cruciaux et de défis technologiques de premier plan qui a nourri ma motivation et ma détermination à entreprendre cette thèse de doctorat. En tant qu'étudiant en informatique, les attaques DDoS m'ont toujours fasciné, non seulement en raison de leur évolution constante, mais aussi en raison des profondes répercussions qu'elles ont dans notre monde interconnecté. La recherche académique offre un cadre d'investigation rigoureux et impartial, qui permet d'explorer en profondeur les problèmes liés à la détection et à la mitigation des attaques DDoS. Elle repose sur une base solide constituée de preuves tangibles et de données empiriques, à la différence des approches basées sur des arguments marketing ou des perspectives partiales, susceptibles de générer des résultats fragiles. Ma thèse s'inscrit pleinement dans cette démarche, visant à enrichir l'état de l'art en développant des approches novatrices pour la détection des attaques DDoS dans le contexte complexe des fournisseurs de services *cloud*.

En résumé, cette thèse s'inscrit dans un contexte où l'informatique occupe une place centrale dans notre société, où les cyberattaques ont atteint un niveau alarmant, et où les attaques DDoS représentent un défi technologique complexe en constante croissance.

L'informatique étant omniprésente, elle exerce une influence indéniable sur notre vie quotidienne, façonnant nos interactions, notre travail, nos loisirs, et même notre vision d'avenir. Face à cette dépendance croissante, les cyberattaques sont devenues une préoccupation majeure, une menace qui mine la sécurité numérique et économique de notre société. Parmi elles, les attaques DDoS se démarquent par leur sophistication et leur impact redoutable, perturbant de manière significative la stabilité des services en ligne que nous tenons pour acquis.

En fin de compte, cette thèse ne se limite pas à une simple étape de ma formation académique, mais constitue un engagement profond envers la société et la sécurité de nos données.

Contexte de recherche

Les recherches que j'ai entreprises s'inscrivent au sein d'une Convention Industrielle de Formation par la Recherche (CIFRE), établie en partenariat entre la société OVHcloud et le Centre de Recherche en Informatique, Signal et Automatique de Lille (CRIStAL) de l'Université de Lille.

OVHcloud, créée en 1999 par Octave Klaba, est une entreprise française dans le domaine du *cloud computing*. Elle offre une gamme complète de services d'héber-

gement web, d'infrastructure *cloud* et de solutions dédiées aux entreprises et aux développeurs. Présente à l'échelle mondiale grâce à ses centres de données répartis à travers le globe. Elle occupe une position de premier plan dans l'industrie du *cloud*, fournissant des solutions fiables et performantes à une clientèle internationale.

Le CRiStAL, au sein de l'Université de Lille, est un établissement de recherche dédié à l'avancement des sciences informatiques, du traitement du signal et de l'automatique. Fort de ses collaborations interdisciplinaires et de ses projets de recherche de pointe, le CRiStAL joue un rôle significatif dans la résolution des défis technologiques contemporains. Il contribue également de manière essentielle à la formation de la prochaine génération de chercheurs et d'ingénieurs, tout en favorisant l'innovation et le transfert de technologie vers l'industrie.

C'est dans ce contexte que j'ai entrepris ces recherches, sous la direction de Gilles Grimaud et Michaël Hauspie, membres de l'équipe de recherche *Extra Small Extra Safe* (2XS). Au sein de la société OVHcloud, j'ai bénéficié de l'encadrement de François Serman, puis de Tristan Groléat et Xavier Guillaume. Je tiens également à souligner la précieuse assistance de nombreuses personnes au sein de ces deux entités pour mener à bien mes travaux.

Structure du mémoire

Après avoir présenté le contexte de mes recherches, la structure de ce manuscrit se décompose de la manière suivante :

Dans le chapitre 1, je propose un état de l'art visant à approfondir la compréhension du contexte technique dans lequel s'inscrivent mes travaux. Je commence par une présentation du *cloud computing,* en mettant en évidence les caractéristiques de l'infrastructure propre aux fournisseurs de services *cloud.* Ensuite, j'explore en détail les attaques DDoS, en exposant les différentes variétés que l'on peut rencontrer. Je présente également une chronologie des attaques DDoS qui ont marqué l'industrie depuis leur apparition. Enfin, je décortique les raisons et spécificités qui font des fournisseurs de services *cloud*[1] des cibles privilégiées pour les attaquants. Cette section se termine par une présentation des différentes stratégies de détection des attaques DDoS volumétriques identifiées dans la littérature.

Le chapitre 2 expose la problématique à laquelle ce manuscrit répond, ainsi que l'approche méthodologique que j'ai adoptée pour y répondre.

1. Dans la suite de ce manuscrit, chaque fois que nous mentionnerons des fournisseurs, il s'agira toujours de fournisseurs de services *cloud*. Pour éviter d'alourdir le texte, nous n'ajouterons plus « de services cloud ».

Le chapitre 3 traite de ma première contribution, présentée sous forme d'un article scientifique à la conférence internationale *IEEE Cloudnet 2023*, intitulé *Scale matters : a Comparative Study of Datasets for DDoS Attack Detection in CSP Infrastructure.* [Boi+23] Après avoir présenté les différents jeux de données largement utilisés par la communauté scientifique pour la détection d'attaques DDoS, j'expose mon choix de métriques pour la modélisation statistique du trafic nominal et des attaques observées sur l'infrastructure d'OVHcloud. Je détaille également la mise en œuvre de la collecte de données de production pour construire le modèle statistique. Enfin, je compare les jeux de données de la littérature avec le trafic réel d'OVHcloud en utilisant les métriques précédemment sélectionnées. Ce chapitre se clôture par une discussion sur les différences d'échelle entre les attaques subies par OVHcloud et celles des jeux de données de la littérature. J'appuie ces conclusions avec un autre article scientifique présenté à la conférence *IEEE CTISC 2022*, intitulé *One Year of DDoS Attacks Against a Cloud Provider : an Overview,* dans lequel je donne un aperçu macroscopique d'une année d'attaques DDoS chez OVHcloud.

Dans le chapitre 4, je présente une preuve de concept d'un générateur de trafic synthétique pour combler une lacune identifiée lors de mes premiers travaux de thèse. Cette lacune concerne le manque de données réalistes sur le trafic nominal et les attaques au sein d'un fournisseur de grande envergure, tel qu'OVHcloud. Après avoir expliqué la mise en œuvre de ce générateur, je compare les données générées avec celles de la production. En conclusion de ce chapitre, je discute de la valeur ajoutée de ce générateur à la fois pour OVHcloud et pour la communauté scientifique.

Le chapitre 5 détaille l'architecture d'un système de détection que j'ai développé et déposé sous la forme d'un brevet industriel. J'explique l'implémentation ainsi que l'articulation des différentes composantes techniques de ce système de détection des attaques DDoS volumétriques à l'échelle d'un fournisseur de services *cloud.*

Enfin, je clôture ce manuscrit par une synthèse de mes contributions et une analyse des résultats obtenus en réponse à la problématique initiale. J'ouvre également la porte à des travaux futurs, à la fois pour approfondir les recherches que j'ai menées et pour envisager des améliorations possibles aux travaux réalisés au cours de cette thèse.

Chapitre 1

État de l'art

Sommaire

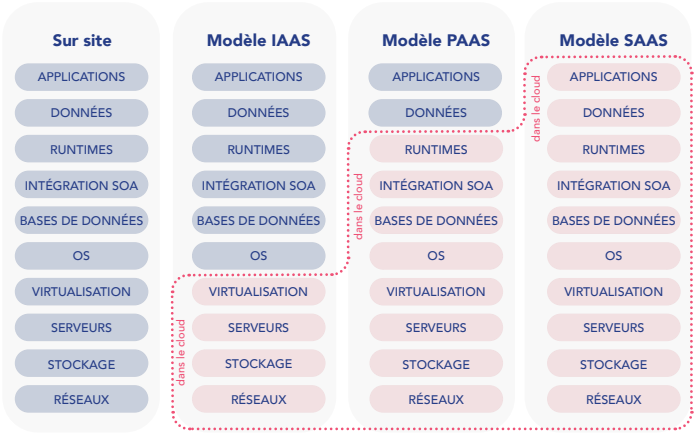

FIGURE 1.1 – Les differents modèles du cloud computing

1.1 Le cloud computing et fournisseurs de services cloud

Le *cloud computing* est désormais un pilier incontournable des infrastructures informatiques modernes, offrant une souplesse, une extensibilité et une accessibilité accrues aux ressources. Les fournisseurs de services *cloud* occupent une place centrale dans la mise à disposition de ces services. Le *cloud computing* représente un modèle qui, comme nous le découvrirons plus en détail par la suite, permet d'accéder à diverses ressources informatiques, telles que des serveurs, du stockage, des bases de données, des logiciels, et bien plus encore, via Internet, de manière à la demande et sans nécessiter la gestion de l'infrastructure matérielle sous-jacente [Fur10].

1.1.1 Modèles de Service

Le *cloud computing* propose une gamme de modèles de services qui répondent aux divers besoins des utilisateurs. Les trois modèles les plus courants sont l'*Infrastructure as a Service* (IaaS), le *Platform as a Service* (PaaS) et le *Software as a Service* (SaaS) [DR17a]. Comme illustré dans la Figure 1.1, chacun de ces modèles offre une approche unique pour l'accès aux ressources et aux logiciels [Al-+13].

Infrastructure as a Service

IaaS constitue la couche la plus fondamentale du *cloud computing*. Elle fournit aux utilisateurs un environnement qui leur permet d'accéder à des ressources matérielles, telles que des serveurs, des espaces de stockage, des réseaux et des systèmes

d'exploitation. Les utilisateurs d'IaaS ont un contrôle total sur ces ressources, ce qui leur permet de déployer des machines virtuelles, de configurer des réseaux, et de gérer leur propre système d'exploitation. IaaS est particulièrement adapté aux entreprises ayant besoin de flexibilité et d'évolutivité pour leurs charges de travail, car il permet d'ajuster rapidement les ressources en fonction des besoins fluctuants [BJJ10].

Platform as a Service

PaaS se situe au niveau supérieur et propose une plateforme complète pour le développement, le déploiement, et la gestion d'applications. Avec PaaS, les développeurs peuvent se concentrer sur la création de logiciels sans se soucier de la gestion de l'infrastructure sous-jacente. Les services PaaS incluent fréquemment des outils de développement, des bases de données, des environnements d'exécution d'applications, et des services de gestion des ressources. Cela accélère le processus de développement et permet aux équipes de se consacrer à l'innovation logicielle, tout en profitant de la scalabilité offerte par le cloud [KR10].

Software as a Service

SaaS représente le modèle de service le plus visible pour de nombreux utilisateurs finaux. Il fournit des applications logicielles entièrement hébergées, accessibles via une interface web. Les utilisateurs n'ont pas besoin d'installer ni de gérer des logiciels localement, car tout fonctionne à partir des serveurs du fournisseur SaaS. Les applications SaaS couvrent un large éventail de domaines, des outils de productivité tels que la suite *Office 365* aux applications métier spécialisées. SaaS offre une simplicité d'utilisation inégalée, nécessite une maintenance minimale, et assure une accessibilité globale, en faisant le choix idéal pour de nombreuses entreprises souhaitant optimiser leurs opérations [Sun+07].

En résumé, ces trois modèles de services, à savoir IaaS, PaaS, et SaaS, proposent des solutions adaptées à divers besoins en informatique et en développement. La compréhension de ces modèles et de leurs avantages peut aider les organisations à prendre des décisions éclairées pour la gestion de leurs ressources informatiques et de leurs applications.

1.1.2 Modèles de Déploiement

Les modèles de déploiement en *cloud computing* déterminent la manière dont les ressources sont déployées et gérées pour répondre aux besoins spécifiques des organisations. Il existe principalement trois modèles de déploiement : le *cloud public*, le *cloud privé* et le *cloud hybride*. Chacun de ces modèles présente ses avantages et

ses inconvénients, et le choix dépend largement des exigences de sécurité, de contrôle et de flexibilité de l'organisation [DR17b].

Cloud Public

Le *cloud public* est l'environnement le plus courant, où les ressources sont partagées entre de multiples clients sur une infrastructure commune gérée par un fournisseur. Cette approche est économique, évolutive et facile à mettre en œuvre, mais elle peut soulever des préoccupations en matière de sécurité et de confidentialité, car les données et les ressources sont partagées [ABA17 ; HW10a ; Li+10 ; Rom12].

Le *cloud public* offre une série d'avantages significatifs pour les entreprises. L'un des principaux avantages est sa capacité à offrir une évolutivité et une flexibilité inégalées. Les fournisseurs de *cloud public* proposent des ressources à la demande, ce qui signifie que les entreprises peuvent facilement augmenter ou réduire leurs capacités en fonction des besoins. Cette flexibilité est particulièrement précieuse pour les charges de travail variables ou saisonnières, permettant de payer uniquement ce qui est utilisé [Put+15].

Un autre avantage clé est la maîtrise des coûts. Contrairement à la gestion d'une infrastructure sur site, où les coûts d'achat, de maintenance et de mise à niveau du matériel peuvent être élevés, le modèle de tarification à l'utilisation des services *cloud publics* permet de réduire les coûts d'exploitation. Les entreprises ne paient que pour les ressources qu'elles consomment, ce qui contribue à une gestion budgétaire plus efficace [SH17].

De plus, la gestion simplifiée est un atout majeur. Les fournisseurs de *cloud public* prennent en charge la gestion de l'infrastructure, y compris la maintenance matérielle et les mises à jour. Cela décharge les équipes internes de tâches fastidieuses et leur permet de se concentrer sur le développement et la gestion des applications et des services [HW10b].

En outre, le *cloud public* offre une disponibilité élevée grâce à la mise en place d'architectures redondantes. Les fournisseurs investissent massivement dans la création de centres de données hautement disponibles, réduisant ainsi les risques de temps d'arrêt imprévus [Put+15].

Enfin, en matière de sécurité, les fournisseurs de *cloud public* mettent en œuvre des mesures de sécurité, telles que la surveillance continue des menaces et la gestion des correctifs de sécurité. Ils fournissent également des outils de sécurité aux clients pour protéger leurs données [RWW12].

Malgré ses avantages, le *cloud public* présente également des inconvénients à prendre en compte.

L'un des principaux inconvénients est la dépendance au fournisseur. En optant pour un *cloud public*, une entreprise dépend des politiques, de la disponibilité des services et de la continuité des opérations du fournisseur. Les interruptions de service ou les modifications unilatérales des conditions peuvent entraîner des perturbations pour l'entreprise cliente [JG+11].

La sécurité et la confidentialité sont également des préoccupations. Bien que les fournisseurs de *cloud public* investissent dans des mesures de sécurité, la gestion de données sensibles dépend également des pratiques de l'entreprise. Les clients doivent être vigilants en matière de protection des données et de conformité réglementaire [Ben+18 ; RWW12].

Les coûts variables sont un autre défi. Bien que le modèle de tarification à l'utilisation puisse réduire les coûts, il peut être difficile de prévoir les dépenses mensuelles en fonction de l'utilisation réelle. Cela peut rendre la budgétisation plus complexe pour certaines entreprises [SH17].

La latence réseau est également un inconvénient potentiel, en particulier pour les applications sensibles au temps. Les performances dans le *cloud public* dépendent de la qualité de la connexion réseau entre l'entreprise et le centre de donnée du fournisseur, ce qui peut entraîner des retards pour certaines applications [Wan10].

Enfin, les services *cloud publics* sont généralement standardisés pour tout les clients, ce qui peut limiter la personnalisation pour répondre aux besoins spécifiques de l'entreprise. Certaines entreprises avec des exigences très particulières peuvent trouver cela restrictif [Fis+13].

Cloud Privé

Le *cloud privé* est une infrastructure dédiée à une seule organisation. Il peut être géré en interne ou par un fournisseur de services, offrant un niveau de contrôle et de sécurité plus élevé par rapport au *cloud public*. Les organisations ayant des exigences strictes en matière de sécurité et de conformité, telles que les institutions financières ou gouvernementales, optent souvent pour un *cloud privé* [ABA17 ; Dua+13 ; XX16].

Le principal avantage du *cloud privé* réside dans sa sécurité renforcée et sa confidentialité. Les organisations bénéficient d'un contrôle total sur leurs données et leurs applications, car elles sont hébergées sur des serveurs dédiés, généralement situés dans le centre de donnée de l'entreprise. Cela garantit une protection des données sensibles et convient particulièrement aux entreprises soumises à des réglementations strictes en matière de confidentialité, telles que la santé ou le paiement.

En outre, le *cloud privé* offre un niveau de contrôle élevé. Les entreprises peuvent personnaliser les configurations et les politiques de sécurité pour répondre à leurs besoins spécifiques. Cette personnalisation permet également une gestion plus fine des ressources, ce qui signifie que les ressources sont utilisées de manière plus efficace.

La performance prévisible est un autre avantage notable. Les ressources d'un *cloud privé* sont dédiées, garantissant ainsi des performances constantes et prévisibles. Cela est essentiel pour les applications critiques qui exigent une réactivité optimale.

L'inconvénient le plus évident du *cloud privé* est son coût élevé. La construction, la maintenance et la gestion d'une infrastructure *cloud privée* nécessitent un investissement considérable en termes de matériel, de logiciels et de personnel qualifié. Les coûts de gestion, de maintenance et de mise à jour peuvent également être significatifs, ce qui rend le *cloud privé* moins abordable pour de nombreuses petites et moyennes entreprises.

De plus, l'évolutivité du *cloud privé* est limitée par la capacité matérielle initiale. L'ajout de ressources supplémentaires peut nécessiter des investissements et des délais importants, ce qui peut être un inconvénient pour les entreprises ayant des besoins évolutifs.

Enfin, comparé aux fournisseurs de *cloud public* qui ont des centres de données répartis dans le monde entier, le cloud privé peut manquer de ressources disponibles dans des emplacements géographiques multiples. Cela peut être un inconvénient pour les entreprises ayant une présence mondiale et nécessitant une disponibilité mondiale des données et des applications.

En résumé, le choix du *cloud privé* dépendra des besoins spécifiques de l'entreprise en matière de sécurité, de performances, de contrôle et de conformité. Bien qu'il offre des avantages indéniables en termes de sécurité et de personnalisation, il est important de peser ces avantages par rapport aux coûts et à la complexité associés à la gestion d'une infrastructure *cloud privée*.

Cloud Hybride

Le *cloud hybride* combine des environnements *cloud publics et privés*, permettant une flexibilité maximale. Il offre la possibilité de déplacer des charges de travail entre les *clouds publics et privés* en fonction des besoins, ce qui est particulièrement utile pour les entreprises dont les besoins fluctuent ou qui ont des exigences de conformité spécifiques [BD18 ; CYC13].

Le *cloud hybride* offre une grande flexibilité pour les entreprises. En combinant des infrastructures *cloud privées et publiques*, il permet aux organisations de gérer leurs charges de travail de manière dynamique. Les ressources peuvent être déployées dans le *cloud public* pour répondre à des pics de demande, tandis que les applications et les données sensibles peuvent être conservées dans un environnement *cloud privé*, offrant ainsi une flexibilité d'évolutivité [Lac11].

La gestion des coûts est un autre avantage important. Les entreprises peuvent

optimiser leurs dépenses en plaçant judicieusement leurs charges de travail dans le *cloud public*, où elles peuvent bénéficier du modèle de tarification à l'utilisation, tout en maintenant le contrôle sur les données sensibles dans le *cloud privé*. Cela permet une utilisation plus efficace des ressources et une meilleure gestion budgétaire.

La sécurité et la conformité sont également renforcées grâce au *cloud hybride*. Les données sensibles peuvent être conservées dans le *cloud privé*, où les entreprises ont un contrôle total sur la sécurité. Cela facilite la conformité avec les réglementations de l'industrie et de la confidentialité des données, ce qui est essentiel pour de nombreuses organisations.

La redondance et la reprise d'activité bénéficient également du *cloud hybride*. Les données peuvent être sauvegardées dans le *cloud public*, offrant une meilleure protection en cas de défaillance du *cloud privé*. Cela garantit une disponibilité élevée des données et une continuité des opérations en cas de catastrophe [Ham19].

Enfin, le *cloud hybride* permet une personnalisation accrue de l'infrastructure. Les entreprises peuvent choisir les solutions *cloud public et privé* qui correspondent le mieux à leurs besoins spécifiques, ce qui les rend plus agiles et adaptées à leurs objectifs [Kui+15].

Mais, la gestion d'un environnement *cloud hybride* peut s'avérer complexe. Il faut gérer deux infrastructures distinctes, ce qui peut nécessiter des compétences et des ressources supplémentaires. La coordination entre le *cloud privé* et le *cloud public* peut être un défi pour les équipes informatiques.

Les coûts d'un *cloud hybride* peuvent être variables et difficiles à prévoir. Les dépenses dépendent des charges de travail déplacées entre les deux environnements, ce qui peut compliquer la budgétisation et la gestion financière [Wei16].

L'intégration des applications et des données entre le *cloud privé* et le *cloud public* peut nécessiter un effort supplémentaire. Une mauvaise intégration peut entraîner des problèmes de compatibilité et de performances, ce qui rend essentiel une planification minutieuse de l'architecture.

La sécurité, bien que renforcée dans le *cloud privé*, nécessite une attention particulière dans un environnement hybride. La gestion de la sécurité peut être complexe, et des erreurs de configuration peuvent entraîner des failles de sécurité.

Enfin, la latence réseau peut être un problème dans un *cloud hybride*, en particulier lorsque des données doivent être transférées entre le *cloud privé* et le *cloud public*. Cela peut affecter les performances de certaines applications, nécessitant une gestion minutieuse de la bande passante et de la connectivité réseau.

En conclusion, le *cloud hybride* offre de nombreux avantages, notamment la flexibilité, la maîtrise des coûts, la sécurité et la personnalisation, mais il nécessite une gestion soignée pour minimiser les inconvénients liés à la complexité, à la budgétisation, à l'intégration, à la sécurité et à la latence réseau. Il est particulièrement adapté aux entreprises ayant des besoins de sécurité stricts, des charges de travail variables

et des exigences de conformité réglementaire.

1.1.3 Conclusion

En conclusion, le *cloud computing* offre un éventail de solutions, chacune avec ses avantages et ses inconvénients distincts. Le *cloud public* se démarque par sa flexibilité, sa gestion simplifiée et sa possibilité de passage à l'échelle, mais il peut entraîner des préoccupations en matière de sécurité et de confidentialité. Le *cloud privé*, quant à lui, garantit un contrôle total, une sécurité renforcée et des performances prévisibles, mais à un coût élevé et avec une complexité de gestion accrue. Le *cloud hybride* offre un équilibre entre les deux mondes, offrant flexibilité, sécurité et personnalisation, mais nécessitant une gestion soignée et pouvant entraîner des défis d'intégration. Par ailleurs, du point de vue d'un fournisseur de services cloud, la gestion de ce modèle peut parfois s'avérer complexe. Pour nous, cela revient à mettre notre cloud privé à la disposition de nos clients. En conséquence, nous devons accorder une attention particulière à la gestion et à l'isolation des données de chaque client, ainsi qu'à la distinction entre les données « internes » et les données « publiques ».

Il est essentiel de souligner que le *cloud hybride*, bien que très pratique pour de nombreux clients, présente des défis en matière de développement continu pour les fournisseurs de services *cloud*. Cela résulte de la nécessité de s'adapter aux évolutions techniques des clients, mais aussi à celles d'autres fournisseurs. Au fil des dernières années, nous avons constaté l'émergence du *multi-cloud* [Hon+19], qui consiste pour un client à utiliser simultanément plusieurs fournisseurs de *cloud public*. Cependant, il n'existe pas encore de normes véritablement appliquées pour réaliser ces interconnexions. Par conséquent, les choix stratégiques d'un fournisseur impactent l'ensemble des fournisseurs qui ont des clients utilisant le *multi-cloud*, créant ainsi une interdépendance entre les prestataires.

En fin de compte, le choix entre ces modèles dépend des besoins spécifiques de l'entreprise, de son budget, de ses exigences de sécurité et de sa tolérance au risque. Le *cloud computing* continue de transformer les infrastructures informatiques, offrant une agilité accrue et une efficacité opérationnelle pour les entreprises du monde entier. Cependant, il est essentiel d'évaluer attentivement les options disponibles pour prendre des décisions éclairées qui répondent aux besoins uniques de chaque organisation.

1.2 L'Infrastructure d'un Fournisseur de Services Cloud

L'infrastructure d'un fournisseur de services *cloud* est le socle technologique qui permet de proposer des services informatiques et de stockage en ligne à des utilisateurs

du monde entier. Cette infrastructure est une composante complexe, conçue pour assurer la disponibilité, la sécurité et la performance des services *cloud*. Dans cette section, nous explorerons les éléments clés de cette infrastructure, notamment les Points de Présence (PoP), la notion de tiers télécom, le cheminement du trafic réseau du client final aux serveurs hébergés dans les centres de donnée, la définition des centres de donnée, des connexions entre serveurs et de la backbone.

1.2.1 Notion de Tiers Télécom

Les réseaux de niveau 1, de niveau 2 et de niveau 3 sont des classifications utilisées pour décrire les rôles et les responsabilités des fournisseurS d'accès à Internet ou FAI (aussi appelé fournisseur de services Internet ou FSI) au sein de l'infrastructure qui fait Internet. Ces niveaux aident à catégoriser les FAI en fonction de leur taille d'infrastructure matérielle, de leur envergure des interconnexions (*peering*) et de l'étendue de leur infrastructure réseau [GMB16].

Les réseaux de niveau 1 se trouvent au sommet de la hiérarchie dans le système de *peering*[1] Internet.

Ils jouissent d'une présence mondiale, avec une infrastructure réseau vaste et étendue sur de nombreux pays et régions à travers le globe. Ce qui distingue les réseaux de niveau 1, par rapport aux réseaux de niveau 2 et de niveau 3, c'est qu'ils ne s'appuient pas sur l'achat de services de transit Internet auprès d'autres FAI. Au lieu de cela, ils établissent des accords de *peering* avec d'autres réseaux de niveau 1, et parfois avec de grands FAI régionaux. Ces réseaux sont responsables de la gestion de leurs propres Systèmes Autonomes (SA).[2]

Historiquement, on les désignait comme des FAI « *free transit* », car ils ne dépendaient pas d'autres réseaux pour acheminer leur trafic. Cependant, il est important de noter que de nos jours, ils établissent des contrats bilatéraux avec d'autres réseaux, ce qui signifie qu'ils participent activement aux arrangements financiers pour l'acheminement du trafic. Cette évolution a modifié la dynamique, et l'expression FAI « *free transit* » ne correspond plus tout à fait à leur fonctionnement actuel.

Ils maintiennent une partie significative de la table de routage mondiale, annonçant leurs préfixes d'adresses IP via le protocole BGP[3] pour assurer un routage efficace

1. Le peering est un accord entre deux FIA pour permettre l'échange direct de trafic Internet, améliorant ainsi la vitesse et l'efficacité des connexions en contournant les réseaux intermédiaires.

2. Les Systèmes Autonomes sont des entités réseau sur Internet qui suivent une seule politique de routage et sont identifiés par un numéro unique, ce qui facilite la gestion et la coordination du trafic sur le réseau.

3. BGP, ou *Border Gateway Protocol*, est un protocole de routage utilisé sur Internet pour aider les routeurs à prendre des décisions sur la meilleure façon de diriger le trafic entre différents réseaux

du trafic Internet. Les réseaux de niveau 1 sont réputés pour leur haute fiabilité et leur disponibilité, étant considérés comme l'épine dorsale d'Internet. Pour les niveaux 1, on considère actuellement cette liste : Orange, Cogent, Deutsche Telekom, Lumen, Arelion, Zayo, Telecom Italia, Liberty Global, Comcast, Telefonica, Verizon, Tata, GTT, PCCW, NTT.

Les réseaux de niveau 2, souvent désignés sous FAI régionaux ou nationaux, se caractérisent par leur présence géographique limitée à une région ou à un pays donné. Ils ont pour mission de fournir une connectivité Internet à des zones géographiques spécifiques ou à des marchés particuliers. Contrairement aux réseaux de niveau 1, les réseaux de niveau 2 ont généralement recours à l'achat de services de transit Internet auprès de fournisseurs de niveau 1 pour établir leur connexion à Internet.

Ces réseaux de niveau 2 ont également la possibilité de conclure des accords de peering avec d'autres réseaux de niveau 2 et parfois avec des fournisseurs de niveau 1 de plus grande envergure, ce qui leur permet d'échanger efficacement du trafic. Ils jouent un rôle fondamental dans l'expansion de l'accès à Internet pour les utilisateurs finaux au sein de leurs régions spécifiques. Pour illustrer, des acteurs comme Hurricane Electric ou Fibrenoire sont des exemples de réseaux de niveau 2 bien connus.

En ce qui concerne OVHcloud, ayant une connexion directe avec plus de 10 des 14 réseaux de niveau 1 mentionnés précédemment, nous disposons d'une connectivité étendue. Cependant, il est important de noter que nous ne revendons pas notre connectivité à des fournisseurs de niveau 3, ce qui complique la catégorisation de notre position en tant que niveau 2. Nos statuts ne correspondent pas strictement aux définitions traditionnelles de niveaux de réseaux.

Les réseaux de niveau 3 sont généralement des FAI plus petits qui desservent des communautés locales ou des groupes de clients spécifiques. Ils ont une empreinte géographique limitée et se concentrent sur la fourniture de services Internet à une zone localisée, telle qu'une ville ou une commune. Les réseaux de niveau 3 achètent souvent des services de transit Internet à des fournisseurs de niveau 1 ou de niveau 2 pour connecter leurs clients à l'Internet plus vaste. Ils peuvent disposer d'une infrastructure et de ressources plus limitées par rapport aux fournisseurs de niveau 1 et de niveau 2. Ces FSI jouent un rôle essentiel dans la fourniture de services Internet aux zones moins densément peuplées et aux marchés de niche.

En résumé, la classification des FAI en réseaux de niveau 1, de niveau 2 et de niveau 3 est basée sur leur envergure, leur couverture géographique et leur rôle au sein de l'écosystème Internet. Les réseaux de niveau 1 sont les plus grands et les plus mondiaux, les réseaux de niveau 2 opèrent à un niveau régional ou national,

autonomes.

tandis que les réseaux de niveau 3 sont plus petits et desservent des marchés locaux ou de niche. Ces niveaux aident à illustrer la nature diversifiée et interconnectée de l'infrastructure Internet.

1.2.2 Points de Présence

Un Point of Presence (PoP) est un élément fondamental de l'infrastructure de réseau. Il s'agit d'un emplacement physique situé stratégiquement dans une région ou une ville. Ce site physique peut être un bâtiment, une installation de télécommunications ou même une armoire de rue, en fonction de la portée et de l'échelle du réseau. Les PoPs sont souvent répartis géographiquement pour assurer une couverture étendue [Zit03].

Au cœur d'un PoP se trouvent des équipements de réseau tels que des routeurs, des commutateurs, des serveurs, des amplificateurs de signal et d'autres dispositifs de transmission de données. Ces équipements sont interconnectés pour acheminer efficacement le trafic Internet ou d'autres types de trafic de données.

Ils constituent la base de notre infrastructure, représentant le point de départ essentiel de notre réseau. Souvent gérés par des prestataires pour le compte de plusieurs hébergeurs, dont nous faisons partie, ces emplacements physiques sont cruciaux pour notre activité. En général, notre facturation est basée sur la quantité d'espace occupée par nos équipements, mesurée en unités de hauteur (U) [4], ainsi que sur la consommation énergétique de ces équipements.

La gestion des PoPs peut s'avérer complexe, en particulier dans le cas de PoPs éloignés, comme celui de Singapour, qui se situe considérablement loin du siège d'OVHcloud. Les interventions nécessaires dans ces emplacements sont coordonnées via des notes de service, un processus qui peut prendre du temps pour le prestataire. C'est pourquoi nous nous efforçons de mettre en place des équipements fiables et de n'utiliser que des technologies éprouvées dans ces emplacements stratégiques.

Cette approche vise à minimiser les interruptions de service et à garantir la continuité de nos opérations, même dans des régions géographiquement éloignées. Elle s'inscrit dans notre engagement envers la qualité et la fiabilité de nos services, ce qui est essentiel pour répondre aux besoins de nos clients à travers le monde.

Les PoPs ne servent pas seulement à relier des réseaux, ils offrent également des services aux clients finaux. Les clients peuvent être des entreprises, des institutions, des fournisseurs de services de *cloud* ou même des particuliers. Les services fournis à partir d'un PoP peuvent inclure l'accès à Internet, la téléphonie IP (VoIP), la vidéoconférence, la diffusion de contenu, et bien plus encore.

4. L'espace en U, dans le contexte du réseau matériel, fait référence à l'espace physique disponible dans une armoire de communication ou un *rack* pour l'installation d'équipements tels que des serveurs, commutateurs, ou autres dispositifs de réseau.

Un aspect essentiel des PoPs est l'optimisation du trafic. Les routeurs et les commutateurs dans un PoP utilisent des algorithmes de routage pour déterminer le chemin le plus efficace pour les données en transit. Cela permet de minimiser la latence, de réduire la congestion et d'assurer une disponibilité élevée [Cha05].

Pour garantir la disponibilité continue des services, les PoPs sont souvent conçus avec des mesures de redondance. Cela signifie qu'il existe des équipements de secours prêts à prendre le relais en cas de panne, assurant ainsi une résilience accrue du réseau.

Les PoPs sont également soumis à des mesures de sécurité rigoureuses. Ils sont sécurisés physiquement et logiquement pour protéger les équipements et les données sensibles qui transitent par eux. Des contrôles d'accès, des pare-feu et d'autres dispositifs de sécurité sont couramment utilisés pour empêcher les intrusions non autorisées.

Enfin, les opérateurs de réseau surveillent en permanence les PoPs pour s'assurer de leur bon fonctionnement. Des outils de gestion et de surveillance permettent de diagnostiquer les problèmes, de réaliser des mises à jour logicielles et de gérer les capacités du réseau. En somme, les PoPs jouent un rôle essentiel dans la mise à disposition de services Internet, de télécommunications et de connectivité à grande échelle, en agissant comme des points de convergence clés au sein de l'infrastructure de réseau.

1.2.3 Centres de données

Un centre de données, au sein d'un fournisseur de *cloud* par exemple, représente l'infrastructure physique essentielle qui sous-tend l'ensemble des services *cloud* offerts aux clients. Ces centres de données sont généralement situés dans des emplacements stratégiques, soigneusement choisis pour garantir la connectivité réseau, la redondance électrique, et se conformer aux réglementations en vigueur. Les centres de donnée sont des installations de grande envergure, pouvant occuper de vastes espaces pour loger une multitude de serveurs, de dispositifs de stockage, de routeurs, de commutateurs, et d'autres équipements informatiques [BHR19].

Les serveurs sont au cœur même d'un centre de donnée. Ils sont conçus pour exécuter diverses applications, héberger des machines virtuelles et stocker d'énormes quantités de données. Dans un centre de donnée typique, on peut trouver des milliers, voire des dizaines de milliers de serveurs, disposés en *racks*[5] et maintenus pour garantir leur disponibilité et leur maintenance aisée [Fur10].

La gestion du stockage de données est une autre composante majeure des datacenters. Ces installations sont équipées de dispositifs de stockage, tels que des disques durs traditionnels, des disques SSD (*Solid-State Drive*) haute performance, et des

5. Les *racks* sont des structures verticales ouvertes ou fermées utilisées dans les centres de données et les salles informatiques pour organiser et monter de manière ordonnée des équipements informatiques tels que serveurs, commutateurs, et dispositifs de stockage.

systèmes de stockage en réseau, comme les NAS (*Network Attached Storage*) et les SAN (*Storage Area Network*). Ces ressources de stockage permettent de gérer la multitude de données générées et stockées par les clients du fournisseur, qu'il s'agisse de stockage d'objets, de fichiers ou de bases de données [Rao+16].

Les centres de donnée sont fortement dépendants d'un réseau de communication haut débit. Ces centres de données sont connectés à un réseau robuste qui assure la communication fluide entre les serveurs, les utilisateurs finaux, et d'autres datacenters ou points de présence. Cette connectivité réseau est essentielle pour fournir des services *cloud* à faible latence et une bande passante élevée [Hab+12].

La sécurité revêt une importance cruciale dans un centre de donnée. Les fournisseurs mettent en œuvre des mesures de sécurité physiques et logiques. Cela inclut des contrôles d'accès stricts, des caméras de surveillance, et souvent la présence de gardes de sécurité pour protéger les installations physiques. Sur le plan informatique, des pare-feu, des systèmes de détection d'intrusion, et des protocoles de gestion des incidents sont en place pour prévenir les cyberattaques et les menaces potentielles [AGM16].

La gestion de l'environnement est également une préoccupation majeure dans un centre de donnée. Ces installations génèrent une chaleur considérable en raison du fonctionnement continu des serveurs. Par conséquent, des systèmes de refroidissement sophistiqués sont déployés pour maintenir une température optimale, tout en économisant l'énergie [Bil+14]. Chez OVHcloud nous fabriquons nos propres serveurs, construisons nos propres centres de données et entretenons des relations solides et à long terme avec d'autres partenaires technologiques, avec un objectif : offrir les solutions les plus innovantes avec le meilleur rapport qualité/prix. L'un des exemples les plus évident, et étroitement lié à notre développement, est l'idée d'utiliser de l'eau pour refroidir nos serveurs [Hna+22].

En somme, un centre de donnée au sein d'un fournisseur de *cloud* est un élément central de l'infrastructure qui permet de fournir des services *cloud* à grande échelle. Il combine des ressources informatiques, une connectivité réseau, une sécurité robuste, une gestion efficace, et une redondance électrique pour répondre aux besoins variés des clients en matière de stockage, de calcul, et de traitement de données, tout en garantissant la disponibilité et la performance continue des services *cloud*.

1.2.4 Interconnexion des Serveurs et backbones

L'interconnexion des serveurs à la *backbone* (épine dorsale) dans le cadre d'un fournisseur de *cloud* est un processus fondamental pour assurer la connectivité, la disponibilité et la performance des services *cloud*. Dans ce contexte, la *backbone* du réseau est une infrastructure centrale à haute capacité qui permet de relier différents emplacements, tels que les centres de donnée, les points de présence (PoPs) et d'autres

nœuds du réseau [DPŞ14 ; Mar+08].

À l'intérieur des centres de donnée, une multitude de serveurs est déployée pour exécuter des applications et gérer les services *cloud*. Ces serveurs sont organisés en *racks* et en clusters, comme nous l'avons vu plutôt, pour une gestion efficace des ressources informatiques. Cependant, pour que ces serveurs puissent communiquer avec le monde extérieur et entre eux, ils doivent être connectés à la *backbone* du réseau [BHR19].

La connectivité réseau locale à l'intérieur d'un centre de donnée est gérée par des commutateurs et des routeurs locaux, assurant le routage interne et la gestion du trafic local. Cependant, pour établir une connexion avec la *backbone* du réseau, des liaisons à haute capacité, telles que des liens en fibre optique, sont mises en place pour relier le centre de donnée aux points d'interconnexion [Wan+15].

Ces points d'interconnexion, parfois appelés « points de *peering* », sont des emplacements stratégiques où plusieurs réseaux différents se rejoignent. Ils jouent un rôle crucial en permettant aux données de passer efficacement d'un réseau à un autre. Cela garantit une connectivité fluide et réduit les temps de latence pour les services *cloud* [Yeg+19].

Pour garantir la disponibilité continue des services *cloud*, les fournisseurs mettent en place des mesures de redondance, ce qui signifie qu'ils ont souvent plusieurs points d'interconnexion vers la *backbone*. Cette redondance assure que, même en cas de panne d'un point d'interconnexion, la connectivité reste intacte. De plus, les capacités de réseau sont conçues pour être évolutives, permettant aux fournisseurs de *cloud* d'ajouter de la capacité en fonction de la demande croissante [KM15].

La gestion de la qualité de service (QoS) est également essentielle. Elle permet de classer le trafic par priorité en fonction de l'importance des applications et des services, assurant ainsi des performances optimales pour les applications sensibles à la latence, comme la voix sur IP ou la vidéoconférence [Zhe+12].

Enfin, la sécurité est un élément clé de l'interconnexion des serveurs à la *backbone*. Les fournisseurs de *cloud* mettent en œuvre des mesures de sécurité strictes pour protéger les données et les services contre les menaces potentielles. [TM11].

En résumé, l'interconnexion des serveurs à la *backbone* du réseau est un aspect essentiel de l'infrastructure d'un fournisseur de *cloud*. Elle assure une connectivité rapide, fiable et sécurisée entre les serveurs dans les datacenters et le reste du monde, contribuant ainsi à fournir des services *cloud* performants et accessibles [Sal+22].

1.3 Les attaques DDoS et DoS

1.3.1 Qu'est-ce qu'une attaque DDoS et DoS

Une attaque par déni de service (DoS), *Denial of Service* en anglais est une attaque DoS est généralement orchestrée par un seul attaquant ou un petit groupe d'attaquants, et elle ne nécessite pas une multitude de *bots*. L'attaquant identifie une vulnérabilité ou une faiblesse dans la cible et exploite cette vulnérabilité en envoyant un volume de trafic massif, parfois faible, de requêtes ou de données malveillantes à la cible. Les ressources de la cible, telles que le processeur, la mémoire et la bande passante, la capacités à ouvrir de nouvelles connexions, sont rapidement sollicitées au-delà de leurs capacités, ce qui peut entraîner une dégradation des performances du système ou une panne complète. Les répercutions de ces attaques sont les mêmes que celles liées aux attaques DDoS [Liu+09].

Une attaque par déni de service distribué (DDoS), *Distributed Denial of Service* en anglais, représente une forme d'attaque informatique. Son objectif principal est de perturber ou d'indisposer un service en ligne, un site web ou un réseau spécifique en submergeant la cible avec une quantité massive de trafic malveillant. La caractéristique principale d'une attaque DDoS est sa nature distribuée, car elle implique un grand nombre d'ordinateurs ou de dispositifs compromis, souvent appelés « *bots* » ou « zombies ». Ces machines infectées sont coordonnées de manière synchronisée par les attaquants pour mener l'attaque [MR04].

Pour mettre en œuvre une attaque DDoS, les attaquants commencent par recruter ces *bots*. Cela se fait généralement en infectant des ordinateurs ou des dispositifs connectés à Internet avec un logiciel malveillant spécialement conçu. Une fois infectés, ces *bots* sont à la disposition des attaquants, qui les contrôlent à distance à l'aide d'un « *botnet controller* » ou d'un « *botnet herder* », un logiciel malveillant de commande et de contrôle [HBK15].

Après avoir recruté un nombre suffisant de *bots*, les attaquants sélectionnent leur cible, qui peut être un serveur web, un service en ligne, ou toute autre infrastructure connectée à Internet. Les motivations pour cibler une entité spécifique peuvent varier, allant des motifs idéologiques à la concurrence en passant par la simple envie de causer des perturbations [Abh+20].

Une fois la cible choisie, les attaquants envoient des instructions aux *bots* pour qu'ils génèrent un flux massif de trafic malveillant vers la cible. Ce trafic peut être constitué de requêtes $HTTP$, de paquets réseau TCP ou d'autres protocoles réseau, selon les vulnérabilités de la cible. L'objectif est de submerger la capacité de traitement de la cible [KDH21].

Les effets d'une attaque DDoS sont significatifs : la cible devient généralement in-

disponible pour les utilisateurs légitimes, car elle est saturée par le trafic malveillant. Les ressources du serveur, telles que le processeur, la mémoire et la bande passante, sont sollicitées au-delà de leurs capacités, entraînant des surcharges, des pannes ou des ralentissements graves. En conséquence, les entreprises ciblées subissent souvent des pertes financières dues à la perte de revenus, aux coûts de mitigation et à la réparation des dommages causés par l'attaque mais aussi une atteinte à l'image de marque de l'entreprise peut avoir lieu [HSM15 ; Som+16 ; SSP12].

En conclusion, les attaques DoS et DDoS sont des menaces sérieuses pour la disponibilité des services en ligne et des réseaux. La différence réside dans la complexité et la dimension distribuée des attaques DDoS, qui les rendent souvent plus difficiles à contrer.

1.3.2 Les botnets

Un *botnet*, abréviation de « *robot network* » en anglais, désigne un ensemble de dispositifs informatiques interconnectés à des fins malveillante, le plus souvent à l'insu de leurs propriétaires légitimes. Ces dispositifs, tels que des ordinateurs, des serveurs, des smartphones ou même des objets connectés, sont compromis, la plupart du temps, en exploitant des failles de sécurités autour des mécanismes d'authentifications, ou des escalades de privilèges. Un individu malveillant, également appelé opérateur de *botnet*, les contrôle à distance pour exécuter des actions automatisées [HBK15].

Le cœur d'un *botnet* réside dans le serveur de commande et de contrôle ($C\&C$), un élément central géré par les opérateurs du *botnet* (aussi appelé *botmasters*. Ce serveur assure la communication entre les *bots* et les opérateurs, permettant à ces derniers de transmettre des ordres aux dispositifs infectés. Les *bots*, quant à eux, sont les machines compromises, chargées de diverses tâches malveillantes, telles que la collecte de données sensibles, l'envoi de spams, les attaques par DDoS ou la propagation de logiciels malveillants supplémentaires [ZM09].

Les *botmasters*, responsables de la gestion et de l'utilisation du *botnet*, sont ceux qui orchestrent ces actions malveillantes. Ils cherchent constamment à infecter de nouveaux dispositifs pour étendre leur réseau. Pour éviter la détection, ils utilisent des tactiques sophistiquées [Dag+07 ; WSZ08].

Les conséquences des *botnets* sont souvent importantes. Ils peuvent causer d'importants problèmes de sécurité, tels que le vol de données sensibles, la perturbation de services en ligne, des pertes financières et des atteintes à la vie privée [FJ09 ; RMG13 ; Yad22]. La lutte contre les *botnets* nécessite la détection des serveurs $C\&C$, la désinfection des dispositifs infectés et la mise en place de mesures de sécurité pour empêcher de futures infections. C'est un effort conjoint impliquant les gouvernements, les entreprises de cybersécurité et les organismes de réglementation pour protéger les

utilisateurs en ligne [Liu+09 ; MDL17 ; Sto+09].

1.3.3 Types d'attaques DDoS

Les attaques d'épuisement de la bande passante

Les attaques d'épuisement de la bande passante, souvent appelées attaques volumétriques, sont une catégorie courante d'attaques DDoS qui ciblent la capacité de bande passante d'une cible spécifique, telle qu'un serveur, un site web ou une infrastructure réseau. L'essence de ces attaques réside dans la saturation de la bande passante de la cible en submergeant celle-ci avec un volume massif de trafic inutile. Les attaquants utilisent généralement un grand nombre de dispositifs [Mao+06 ; San+15a].

Ces attaques peuvent utiliser divers types de trafic pour saturer la bande passante, notamment des paquets UDP, $ICMP$, TCP ou même $HTTP$, en fonction de la vulnérabilité spécifique qu'ils ciblent. Une caractéristique notable de certaines attaques est l'utilisation d'amplificateurs, tels que des serveurs DNS mal configurés ou des serveurs NTP (*Network Time Protocol*) malveillants, qui renvoient un volume beaucoup plus élevé de données que ce qui a été initialement envoyé, augmentant ainsi la puissance de l'attaque [Ali+16 ; Mia+15 ; RI15].

Les attaquants déploient une variété de techniques pour compliquer la détection et la défense contre leurs attaques. Cela inclut la variation des adresses IP source [6], *IP Spoofing* en anglais, la modification de la taille des paquets et la distribution du trafic depuis plusieurs points d'origine. L'objectif ultime est de provoquer une congestion du réseau chez la cible, ce qui peut entraîner un déni de service. Les conséquences peuvent aller de la lenteur du réseau à des temps d'arrêt complets, en passant par la perturbation des services en ligne et la frustration des utilisateurs légitimes [Sac+10].

Les attaques d'épuisement des ressources

Les attaques d'épuisement des ressources visant à épuiser les ressources critiques d'un système informatique, telles que la capacité de traitement CPU, la mémoire, ou d'autres ressources spécifiques. L'objectif principal de ces attaques est de rendre un service indisponible en forçant la cible à épuiser toutes ses ressources nécessaires pour fonctionner correctement. Ces attaques exploitent les limites des ressources disponibles sur un système et peuvent être dévastatrices si elles réussissent [Pas+17].

Les ressources ciblées dans ces attaques peuvent varier en fonction de l'objectif de l'attaquant. Par exemple, une attaque peut cibler la capacité CPU en envoyant

6. L'IP spoofing est une technique où un expéditeur falsifie délibérément l'adresse IP source d'un paquet réseau pour tromper les destinataires sur l'origine réelle du paquet, souvent utilisée à des fins malveillantes ou pour des activités illicites en ligne.

une grande quantité de demandes de calcul intensif, ou elle peut cibler la mémoire en saturant le système avec des requêtes ou des données volumineuses.

L'objectif ultime de ces attaques est de délibérément causer la saturation ou l'épuisement total des ressources ciblées. Lorsque ces ressources essentielles sont épui-sées, le système visé peut devenir non fonctionnel, entraînant ainsi des temps d'arrêt et un déni de service pour les utilisateurs légitimes. Ce type d'attaque reste possible même en présence de systèmes économiques et évolutifs, car les attaques par épuisement de ressources sont réalisables en augmentant simplement le volume des requêtes de manière à dépasser la capacité de mise à l'échelle du centre de données. Cette réalisation est toutefois plus difficile, voire impossible, sans la distribution de l'attaque sur un botnet (d'où le « D » de DDoS, pour « Distributed Denial of Service »). Les conséquences de telles attaques peuvent être extrêmement graves, se traduisant par des perturbations des opérations commerciales, des pertes de productivité et des coûts associés à la remise en état du système. En conséquence, il est essentiel de mettre en place des mesures de sécurité solides pour détecter et atténuer ces attaques, ainsi que pour prévenir leur impact sur les opérations de l'entreprise. [DD15].

Les attaques applicatives

Les attaques de couches d'application, également connues sous le nom d'attaques de la couche 7 dans le modèle OSI[7], ciblent directement les applications web et les services en ligne en exploitant les vulnérabilités au niveau de l'application elle-même. La couche 7 est la couche supérieure du modèle OSI, responsable de l'interaction entre les applications logicielles et les utilisateurs. Elle englobe des protocoles de communication tels que $HTTP$ pour les sites web, $SMTP$ pour les e-mails, et bien d'autres [BR07 ; PT18].

L'objectif principal de ces attaques est toujours le même, de perturber ou de rendre indisponibles des services en ligne en exploitant les failles au niveau de l'application. Cela peut se produire de plusieurs manières, notamment en surchargeant les serveurs cibles avec un grand nombre de requêtes légitimes, en exploitant des vulnérabilités logicielles connues ou *zero-days*[8], ou en épuisant les ressources du serveur, comme le CPU ou la mémoire, en sollicitant des fonctionnalités gourmandes en ressources [Sho+18 ; SU14].

Les techniques utilisées pour mener ces attaques sont variées et sophistiquées.

7. Le modèle OSI (Open Systems Interconnection) est un cadre de référence qui divise les fonctions de communication réseau en sept couches distinctes, permettant ainsi de comprendre et de concevoir des réseaux informatiques de manière modulaire et interopérable.

8. Une vulnérabilité « zero-day » désigne une faille de sécurité informatique qui est exploitée par des attaquants avant même que les développeurs aient eu la possibilité de la corriger ou de publier un correctif, laissant ainsi les systèmes vulnérables à des attaques.

Parmi elles, on trouve les attaques par inondation, où un grand nombre de requêtes légitimes sont envoyées pour saturer le serveur, les attaques de type *Slowloris* [Sho+18] qui maintiennent de nombreuses connexions ouvertes simultanément en envoyant des requêtes $HTTP$ incomplètes, ainsi que l'injection de code malveillant pour exploiter des vulnérabilités et compromettre le serveur.

1.3.4 Historique des attaques DDoS

Années 1980-1990

Pendant les années 1980 et 1990, les attaques DDoS ont émergé à une époque où l'Internet était encore à ses balbutiements. À leurs débuts, ces attaques étaient relativement simples, menées par des individus malveillants inondant un serveur ou un système cible de requêtes pour le submerger. Le terme « DDoS » a été popularisé dans les années 1990 pour décrire ces attaques, qui impliquaient souvent l'utilisation de *botnets* simples, pour coordonner et amplifier les assauts. Des attaques notables, comme celle contre le fournisseur d'accès Internet Panix en 1996 ont mis en évidence la vulnérabilité des infrastructures Internet de l'époque [Bro+21 ; Men97 ; Naz08].

Années 1990-2000

Entre les années 1990 et 2000, les attaques DDoS ont connu une évolution significative, devenant une préoccupation majeure pour la sécurité en ligne. Durant cette décennie, les attaquants ont accru la puissance de leurs assauts en exploitant des techniques d'amplification, notamment en ciblant des serveurs DNS et NTP mal configurés pour inonder leurs cibles de trafic. Les attaques contre Yahoo et Amazon en 2000 a attiré l'attention sur le potentiel destructeur de ces attaques, tandis que de nouveaux outils, tels que TFN2K et Stacheldraht, sont apparus pour coordonner les attaques de manière plus sophistiquée. Les entreprises ont renforcé leurs défenses avec des pare-feu, des systèmes de détection d'intrusions et des services de protection DDoS professionnels, tandis que les attaquants ont continué à évoluer, ciblant davantage les couches d'application. Cette période a également vu une coopération internationale accrue pour faire face à ces menaces transfrontalières, soulignant l'importance croissante de la sécurité en ligne à l'échelle mondiale [Dit99 ; DL00 ; Ell00 ; GW00 ; Kes00].

Années 2000-2010

Pendant la période allant des années 2000 aux années 2010, les attaques DDoS ont connu une évolution significative et une escalade notable. Les attaquants ont accru

la puissance de leurs assauts en utilisant d'énormes *botnets*, résultant en des interruptions de service prolongées pour les cibles visées. Des *botnets* spécialisés, comme Storm Worm et Conficker, ont émergé pour mener des attaques de grande envergure, tandis que les attaquants ont continué à exploiter des vulnérabilités dans les protocoles DNS pour des attaques par amplification. Les cibles des attaques sont devenues plus variées, allant des sites web aux infrastructures critiques, et les attaques ont été ciblées de manière plus précise. De plus, l'utilisation d'attaques de couches d'application supérieures, telles que $HTTP$ et SSL, a rendu les défenses plus complexes. Pour faire face à cette menace croissante, de nombreuses entreprises ont opté pour des services de protection DDoS professionnels, tandis que les autorités ont intensifié leurs efforts pour poursuivre les auteurs d'attaques DDoS, soulignant ainsi l'importance de la sécurité numérique dans un monde de plus en plus connecté [Hol+08 ; SG10 ; WLZ10].

Années 2010-2020

Entre les années 2010 et 2020, les attaques DDoS ont évolué en devenant plus massives, sophistiquées et diversifiées dans leurs objectifs. Les *botnets* IoT ont dominé le paysage, exploitant la faible sécurité des dispositifs connectés pour lancer des attaques d'une ampleur sans précédent, paralysant des services majeurs tels que ceux de Dyn en 2016. Ces attaques ont également été utilisées à des fins d'extorsion, les cybercriminels menaçant de perturber les services en ligne à moins que des rançons ne soient versées. En parallèle, des groupes d'activistes et des acteurs politiques ont utilisé les DDoS pour faire passer leurs messages et exprimer leurs protestations. La défense en temps réel est devenue essentielle pour contrer ces assauts, avec des services de protection DDoS gérés de plus en plus répandus. La réglementation sur la cybersécurité s'est également renforcée pour dissuader les attaquants. Dans l'ensemble, cette période a souligné la nécessité croissante de se prémunir contre les attaques DDoS, en renforçant les défenses et en renforçant la collaboration internationale pour lutter contre cette menace persistante [Ant+17 ; Bre12 ; DC19 ; HKP13 ; LH21 ; New19 ; Sau14].

Années 2020-2023

Au cours des dernières années, les attaques ont continué à évoluer de manière significative. Nous avons assisté à une montée en puissance spectaculaire des DDoS volumétriques, avec des niveaux de trafic dépassant régulièrement le téraoctet par seconde. Ces attaques massives ont pris pour cible une variété d'organisations, des grandes entreprises aux fournisseurs, perturbant leurs activités en ligne et provoquant des pertes financières considérables. La complexité des attaques a également aug-

menté, avec une utilisation accrue d'attaques « verticales » qui ciblent spécifiquement les couches d'application, rendant la détection et la mitigation plus difficiles [MS21 ; NKD22].

L'offre d'attaques DDoS en tant que service (DDoSaaS) s'est développée, rendant ces attaques accessibles à un public plus large, y compris à des acteurs non techniques. Cette commercialisation a conduit à une prolifération des attaques DDoS et à une complexité croissante de la menace [Kop+19].

Parallèlement, le ciblage sectoriel est devenu plus prononcé, avec des attaques DDoS visant des secteurs sensibles tels que la santé et les infrastructures critiques, suscitant des inquiétudes quant à la sécurité nationale [Rao+20 ; RMD22].

Enfin, la législation sur la cybersécurité s'est renforcée, avec une coopération internationale plus étroite visant à identifier et à poursuivre les attaquants DDoS. La lutte contre ces attaques est devenue un effort global pour garantir la sécurité des services en ligne et la stabilité de l'Internet.

1.4 Gestion des attaques DDoS par les opérateurs du cloud

Les attaques DDoS posent d'importants défis aux fournisseurs de services *cloud*. L'une des principales difficultés réside dans la volumétrie accrue du trafic généré par ces attaques. Les attaquants cherchent à submerger les infrastructures de serveurs et les canaux de communication avec un flux massif de données, rendant les services *cloud* inaccessibles pour les utilisateurs légitimes. La gestion de cette augmentation soudaine du trafic est un défi majeur, nécessitant une capacité de bande passante importante et des ressources de calcul considérables.

Un autre défi majeur réside dans la variété des attaques DDoS. Les attaquants peuvent utiliser différentes techniques, telles que les attaques *UDP Flood*, *SYN Flood* ou *HTTP Flood*, etc, pour cibler diverses couches du réseau ou des applications. Les fournisseurs doivent être préparés à faire face à une gamme variée d'attaques, chacune nécessitant des contre-mesures spécifiques.

Les sources d'attaques DDoS sont également très diversifiées. Les attaquants exploitent, comme nous l'avons décrit des *botnets* réparties dans le monde entier. Cette multiplicité des sources d'attaques rend la détection et la mitigation d'autant plus difficiles, car elles proviennent d'adresses IP différentes et peuvent être coordonnées de manière complexe.

La détection des attaques DDoS représente un autre défi. Identifier une attaque parmi le trafic légitime peut être complexe, car les attaquants cherchent à camoufler leurs activités malveillantes. De plus, même une attaque de plusieurs téraoctets, ne

représente qu'une faible proportion du volume totale de traffic qu'observe un fournisseur de service *cloud*.

La traçabilité des attaques DDoS est souvent limitée. Les attaquants utilisent des techniques pour masquer leurs empreintes, l'*IP spoofing*, rendant difficile l'identification de la source réelle de l'attaque. Cette absence de traçabilité peut compliquer les tentatives de poursuite judiciaire contre les attaquants, ce qui renforce leur impunité.

En outre, les fournisseurs sont confrontés à la pression constante de maintenir la performance et la disponibilité de leurs services, car les clients s'attendent à une accessibilité constante. Les attaques DDoS peuvent perturber la disponibilité, ce qui peut nuire à la réputation du fournisseur et entraîner des pertes financières.

Enfin, les attaques DDoS évoluent constamment, avec de nouvelles tactiques et techniques développées par les attaquants. Les fournisseurs doivent rester constamment à jour avec les dernières menaces et améliorer en permanence leurs stratégies de défense pour faire face à cette évolution rapide des attaques.

En résumé, les fournisseurs doivent faire face à des défis complexes pour atténuer les attaques DDoS, nécessitant des ressources techniques, financières et humaines importantes, ainsi qu'une collaboration étroite avec d'autres acteurs de la sécurité pour maintenir la stabilité et la sécurité de leurs services.

1.4.1 Détection basée sur la signature

La détection d'attaques DDoS basée sur la signature est une méthode de sécurité qui repose sur l'identification de modèles de trafic ou de comportements caractéristiques d'attaques DDoS déjà connues. Les experts en sécurité élaborent des signatures spécifi-ques pour chaque type d'attaque, décrivant les caractéristiques distinctives qui se produisent lors d'une attaque DDoS particulière. Ces signatures peuvent être basées sur des adresses IP sources, des ports, des protocoles, ou des motifs de requêtes spécifiques [DPM21 ; GS14 ; MK20].

Lorsque cette méthode est en place, les systèmes de détection surveillent en permanence le trafic entrant et le comparent aux signatures préalablement définies. Si le trafic correspond à l'une des signatures, l'attaque DDoS est détectée, et une alerte est généralement générée pour informer les administrateurs de sécurité. Dans certains cas, des actions de mitigation, telles que le blocage du trafic malveillant, peuvent être déclenchées automatiquement [AR12].

Les avantages de la détection basée sur la signature résident dans son efficacité pour détecter rapidement les attaques DDoS bien connues. Elle a également tendance à générer moins de faux positifs que certaines autres méthodes, car elle se concentre sur des schémas de trafic déjà identifiés. De plus, sa mise en œuvre est relativement simple, car elle ne nécessite que la configuration des signatures pour les attaques DDoS connues [HS14].

Cependant, cette méthode présente des limites importantes. Elle est inefficace pour détecter de nouvelles attaques DDoS pour lesquelles il n'existe pas encore de signature. Les attaquants peuvent également facilement contourner cette méthode en modifiant légèrement leur attaque pour échapper à la détection basée sur la signature, obligeant ainsi les équipes de sécurité à mettre constamment à jour les signatures. Certaines attaques peuvent même utiliser des techniques d'évasion comme, par exemple, la variation des motifs de trafic, le chiffrement de la charge utile, pour dissimuler leurs actions, rendant la détection encore plus difficile [HS14].

Pour les fournisseurs à grande échelle, ces limitations sont exacerbées. En raison de la variété de leur clientèle et du volume massif de trafic qu'ils gèrent, la création de signatures pour chaque type d'attaque devient difficile et la gestion des signatures existantes devient une tâche complexe. De plus, les attaques DDoS évoluent rapidement, ce qui signifie que de nouvelles attaques peuvent émerger plus rapidement que les signatures ne peuvent être mises à jour, mettant ainsi en danger la stabilité des services.

1.4.2 Détection basée sur le comportement

La détection d'attaques DDoS peut s'appuyer sur différentes approches de sécurité, dont la détection basée sur le comportement [LL05 ; LMN19 ; ZLZ10], l'analyse de la corrélation [JD17 ; Xia+15], et l'apprentissage automatique et l'IA [Abd+22 ; AH18 ; ZZY17]. La détection basée sur le comportement analyse les modèles de trafic et les comportements réseau pour repérer les anomalies, en se focalisant sur des déviations par rapport aux modèles normaux établis à partir de données en temps réel. Lorsqu'une anomalie est détectée, le système génère une alerte, déclenchant éventuellement des actions de mitigation.

L'approche de détection basée sur la corrélation repose sur la consolidation de multiples sources d'informations afin d'identifier des relations ou corrélations entre les diverses données du trafic réseau. Cette méthode utilise des modèles de corrélation pour décrire les schémas normaux de relations entre les différentes données, comme la corrélation entre le trafic entrant et sortant ou les modèles typiques de charge de travail. L'objectif est de créer une vue globale du trafic réseau en prenant en compte diverses sources, telles que les journaux de serveur et les informations de télémétrie réseau.

Pour détecter toute déviation significative en temps réel, les systèmes de détection basés sur la corrélation comparent constamment les données actuelles avec les modèles de corrélation établis. Si une anomalie est repérée, le système génère une alerte pour informer les administrateurs de sécurité, pouvant déclencher des actions de mitigation en fonction de la gravité de l'incident. Cette approche est particulièrement efficace pour détecter des attaques DDoS sophistiquées qui pourraient échapper à des

méthodes de détection plus simples.

Cependant, la mise en place de cette méthode est complexe et exige une expertise approfondie en sécurité réseau. La configuration des systèmes de détection basés sur la corrélation nécessite une attention particulière, et une surveillance continue des données en temps réel est cruciale pour assurer une détection précise. Ces défis sont amplifiés pour les fournisseurs de services cloud en raison du volume massif de données à gérer, nécessitant des ressources informatiques puissantes et une infrastructure réseau robuste. De plus, la variabilité des clients rend difficile l'établissement de modèles de corrélation précis, ajoutant une couche de complexité à la configuration de cette méthode de détection. L'évolutivité de la détection basée sur la corrélation pour répondre aux besoins des fournisseurs peut également représenter un défi majeur en termes d'ingénierie et de ressources.

La détection basée sur l'apprentissage automatique et l'IA représente une approche avancée de la sécurité qui se distingue par l'utilisation de modèles prédictifs formés sur des données historiques pour identifier les comportements anormaux du trafic réseau. Le processus démarre par la collecte minutieuse de données historiques, comprenant des informations sur le volume de données, la fréquence des connexions, les types de protocoles utilisés, etc. Ces données servent à entraîner un modèle d'apprentissage automatique ou d'IA, permettant au modèle de distinguer les comportements normaux des comportements anormaux.

Une fois le modèle entraîné, il est déployé pour surveiller le trafic en temps réel. Continuellement, le modèle analyse les métriques actuelles du trafic, les comparant aux modèles qu'il a appris. En cas d'identification d'anomalies significatives ou de comportements potentiellement malveillants, le modèle génère des alertes destinées aux administrateurs de sécurité. En fonction de la gravité de l'anomalie, des actions de mitigation peuvent être déclenchées automatiquement pour atténuer l'impact de l'attaque.

Bien que cette approche offre la capacité de détecter des attaques DDoS inconnues ou émergentes, elle présente des défis importants. La qualité des données historiques est cruciale pour former des modèles précis, ce qui peut parfois être difficile à obtenir. De plus, la mise en place de systèmes d'apprentissage automatique et d'IA est complexe et peut être coûteuse en termes de ressources informatiques et de compétences en matière de données. Un autre défi réside dans la compréhension des modèles par les experts en IA, car les modèles plus complexes, tels que les réseaux de neurones profonds, peuvent être difficiles à interpréter.

Pour les fournisseurs, la gestion d'énormes volumes de trafic rend la collecte, le stockage et le traitement des données d'apprentissage automatique et d'IA plus complexes. La variabilité des clients nécessite également des ajustements constants des modèles pour tenir compte des différents comportements et profils d'utilisation. En outre, il existe un compromis entre la précision d'un modèle d'IA et son explicabilité.

Les fournisseurs doivent équilibrer ces aspects pour fournir des services de sécurité efficaces tout en expliquant leurs décisions aux clients. En résumé, bien que cette méthode soit puissante, elle comporte des défis liés à la qualité des données, à la complexité de mise en place et à la nécessité de trouver un équilibre entre précision et explicabilité.

1.4.3 Détection basée sur le filtrage IP et la géolocalisation

La détection d'attaques DDoS basée sur le filtrage IP et la géolocalisation est une stratégie de sécurité qui repose sur la gestion du trafic en fonction des adresses IP sources et de la localisation géographique de l'origine du trafic. Cette approche consiste à collecter des informations sur les adresses IP associées à des attaques DDoS connues et à utiliser ces données pour bloquer ou limiter le trafic suspect dès son origine. De plus, elle intègre des informations géographiques pour identifier et agir contre le trafic provenant de régions géographiques suspectes [Rod15 ; XKA16].

L'opération de base de cette méthode est de collecter des informations sur les adresses IP à partir de listes noires d'adresses IP connues pour être associées à des attaques DDoS antérieures. Lorsque le trafic est détecté en provenance de ces adresses IP, il est filtré ou bloqué avant d'atteindre la cible, contribuant ainsi à réduire la charge sur les ressources réseau et les serveurs cibles.

En parallèle, la détection basée sur la géolocalisation s'appuie sur des services de géolocalisation pour identifier l'origine géographique du trafic. Si le trafic provient de régions géographiques inhabituelles ou suspectes, il peut être soumis à des mesures de blocage ou de limitation. Cela peut être particulièrement utile pour contrer les attaques DDoS provenant de zones géographiques connues pour abriter des attaquants.

Les avantages de cette méthode résident dans sa capacité à réduire la charge sur les ressources du réseau et à prévenir les attaques DDoS en bloquant ou en limitant le trafic suspect dès son origine. Elle offre une protection préventive contre les attaques DDoS connues, ce qui peut minimiser leur impact sur les services en ligne. De plus, elle peut être ajustée en temps réel pour répondre à des menaces en constante évolution.

Cependant, cette méthode présente également des limites. Les attaquants peuvent contourner cette approche en utilisant des techniques pour masquer ou modifier leur adresse IP source, rendant ainsi le filtrage IP moins efficace. De plus, le filtrage basé sur la géolocalisation peut entraîner des faux positifs, car certaines adresses IP peuvent être associées à des proxy ou à des VPN, masquant ainsi leur véritable origine. De plus, la géolocalisation n'est pas toujours précise à 100%, ce qui peut entraîner des erreurs de blocage ou de limitation de trafic légitime.

Pour les fournisseurs, les défis sont liés à la gestion de la variété des clients, à l'évolutivité pour gérer un volume massif de trafic, ainsi qu'au contournement potentiel de cette méthode par des attaquants qui utilisent des réseaux de distribution de

contenu (CDN) pour masquer la véritable origine de leur trafic. Par conséquent, cette méthode doit être complétée par d'autres approches de détection et de mitigation pour garantir une sécurité efficace contre les attaques DDoS à grande échelle.

1.4.4 Détection basée sur la détection d'anomalies statistiques

La détection d'attaques DDoS basée sur la détection d'anomalies statistiques est une méthode de sécurité qui se concentre sur la surveillance du trafic réseau pour identifier des comportements inhabituels ou des modèles de trafic atypiques. Cette approche repose sur la création de modèles statistiques du trafic réseau normal et la détection de divergences significatives par rapport à ces modèles [Fei+03 ; Gir+15 ; Kha+18].

Pour commencer, un modèle statistique du trafic réseau normal est établi en analysant les données historiques. Ce modèle prend en compte diverses métriques, telles que le volume de trafic, la fréquence des requêtes, la répartition des ports, etc. Une fois que le modèle statistique est en place, le trafic réseau en temps réel est surveillé en permanence. Les métriques du trafic actuel sont comparées aux valeurs prévues par le modèle statistique.

Si les métriques du trafic actuel dévient de manière significative par rapport aux valeurs prévues par le modèle statistique, cela est considéré comme une anomalie. Ces anomalies peuvent être des pics soudains de trafic, une utilisation excessive de ressources, des connexions suspectes, etc. Lorsqu'une anomalie est détectée, le système génère généralement une alerte pour informer les administrateurs de sécurité. En fonction de la gravité de l'anomalie, des actions de mitigation peuvent être déclenchées automatiquement.

Les avantages de cette méthode résident dans sa capacité à détecter des attaques DDoS inconnues ou émergentes, car elle ne dépend pas de signatures préalablement définies. Elle peut identifier tout comportement anormal, même si cela n'a jamais été observé auparavant. De plus, les modèles statistiques peuvent être ajustés au fil du temps pour prendre en compte l'évolution des modèles de trafic et des comportements malveillants. En outre, cette méthode a tendance à générer moins de faux positifs que certaines autres approches.

Cependant, il y a des limites à considérer. La création de modèles statistiques précis nécessite une phase d'apprentissage initiale, ce qui signifie que la méthode peut prendre un certain temps pour devenir pleinement efficace. De plus, l'analyse statistique du trafic réseau peut être complexe, nécessitant des ressources informatiques considérables pour traiter de grandes quantités de données en temps réel. La configuration des systèmes de détection basés sur la détection d'anomalies statistiques doit être soigneusement effectuée pour éviter de générer un grand nombre de fausses alertes ou pour ne pas manquer de véritables attaques.

Pour les fournisseurs, il existe des défis supplémentaires, notamment la gestion de la variabilité des clients et l'évolutivité des systèmes. En raison de la diversité des clients, les modèles statistiques doivent être adaptés pour tenir compte des différentes configurations et des comportements uniques de chaque client. La gestion de plusieurs modèles pour chaque client peut devenir complexe à grande échelle.

En résumé, la détection basée sur la détection d'anomalies statistiques offre une approche puissante pour la détection des attaques DDoS, en particulier pour les attaques inconnues, mais elle comporte des défis importants en termes d'apprentissage initial, de complexité et de configuration. Les fournisseurs de services *cloud* à grande échelle doivent prendre en compte ces limites et envisager d'autres méthodes de détection et de mitigation pour garantir une sécurité réseau complète.

1.4.5 Détection basée sur la protection cloud

La détection d'attaques DDoS basée sur la protection cloud, *Scrubbing center* en anglais, est une approche de sécurité qui externalise la gestion de la sécurité des réseaux en cas d'attaques DDoS vers des fournisseurs spécialisés. Cette méthode fonctionne en redirigeant le trafic entrant vers l'infrastructure *cloud* de sécurité du fournisseur avant qu'il n'atteigne la destination finale. Une fois dans le cloud de sécurité, le trafic est analysé en utilisant des méthodes de l'état de l'art. Si une attaque est détectée, des mesures de filtrage sont généralement appliquées pour bloquer le trafic malveillant, préservant ainsi le trafic légitime [Bha+16 ; Mou+20 ; Sha+15].

Cette approche présente plusieurs avantages, notamment l'évolutivité, car les fournisseurs spécialisés disposent d'infrastructures massives et de capacités de traitement importantes pour faire face à des attaques de grande envergure. De plus, ils bénéficient d'une expertise spécialisée dans la détection et la mitigation des attaques DDoS, ce qui peut être un avantage précieux pour les organisations qui ne disposent pas de ces compétences en interne. De plus, en redirigeant le trafic vers le *cloud* de sécurité avant d'atteindre la destination finale, cette méthode peut augmenter la latence pour le trafic légitime, réduisant ainsi la qualité des services.

Cependant, il existe également des limites à cette approche. Les organisations qui utilisent la protection *cloud* dépendent de tiers pour la protection de leur infrastructure, ce qui peut soulever des préoccupations en matière de confidentialité et de contrôle. De plus, l'utilisation de services de protection cloud peut entraîner des coûts significatifs, en particulier pour les attaques DDoS de grande ampleur. Enfin, il existe un potentiel de faux positifs, c'est-à-dire que les services de protection *cloud* peuvent parfois marquer à tort du trafic légitime comme malveillant, ce qui peut entraîner la perte de données importantes.

Dans le cas des fournisseurs à grande échelle, la détection basée sur la protection

cloud peut ne pas être pertinente pour plusieurs raisons. Tout d'abord, cela implique que les données sensibles des clients d'un fournisseur de services *cloud* transitent vers un autre via Internet, ce qui est inenvisageable pour la plupart des fournisseurs de services *cloud*.

L'exportation de données clients vers un tiers peut poser divers problèmes par rapport au respect du *Cloud Act* (*Clarifying Lawful Overseas Use of Data Act*) aux États-Unis et du RGPD en Europe. Tout d'abord, il y a des préoccupations liées à la juridiction et à la souveraineté des données. Le *Cloud Act* permet aux autorités américaines d'accéder aux données stockées par des entreprises américaines, même si ces données sont situées en dehors des États-Unis. Cela soulève des questions quant à la protection des données et à la conformité aux lois de protection des données d'autres juridictions. Si des données sont exportées vers un tiers basé aux États-Unis, elles pourraient être soumises à la juridiction américaine, ce qui pourrait compromettre la confidentialité des données et la conformité au RGPD.

En outre, le consentement des utilisateurs est un aspect crucial du RGPD. Nous devons obtenir un consentement explicite des utilisateurs pour collecter, stocker et traiter leurs données personnelles. Lorsque des données sont exportées vers un tiers, il devient essentiel de garantir que les utilisateurs ont donné leur consentement en toute connaissance de cause pour cette action, ce qui peut s'avérer complexe à gérer.

La sécurité des données est également une préoccupation majeure lors de l'exportation de données vers un tiers. Nous devons assurer que le tiers destinataire dispose de mesures de sécurité adéquates pour protéger les données des clients, conformément aux exigences du RGPD. En cas de violation de données, le RGPD impose des obligations strictes en matière de notification des incidents. Si des données clientes sont exportées vers un tiers et qu'une violation survient, il peut être difficile de répondre rapidement et de manière appropriée aux obligations de notification, ce qui pourrait entraîner des sanctions.

De plus, il faut tenir compte de la question de la responsabilité conjointe. Selon le RGPD, lorsque des données personnelles sont partagées avec un tiers, l'entreprise d'origine peut être considérée comme responsable conjointe des données. Cette responsabilité partagée signifie que l'entreprise partage la responsabilité légale en matière de protection des données avec le tiers. Il est donc essentiel d'établir clairement les responsabilités et les engagements contractuels pour garantir la conformité.

Finalement, étant donné le volume massif de trafic ainsi que la fréquence quotidienne d'attaques de grande envergure, plusieurs centaines de téraoctets de trafic malveillant devraient transiter par Internet, engendrant des coûts de fonctionnement ainsi que des risques pour la sécurité des infrastructures prohibitifs.

1.4.6 Conclusion

En conclusion, cette section a exploré en profondeur diverses méthodes de détection d'attaques DDoS, en mettant en évidence leurs avantages, leurs limites et leurs applications spécifiques, notamment dans le contexte des fournisseurs à grande échelle. Nous avons examiné des approches allant de la détection basée sur les signatures à l'utilisation de l'apprentissage automatique et de l'intelligence artificielle, en passant par la détection basée sur la détection d'anomalies statistiques, ainsi que la protection *cloud*.

Chacune de ces méthodes présente ses propres forces et faiblesses, et le choix de la méthode appropriée dépendra des besoins spécifiques de l'organisation, de son infrastructure, de ses ressources et de ses compétences en matière de sécurité. Les signatures sont efficaces pour détecter des attaques connues, tandis que l'apprentissage automatique offre la possibilité de détecter des attaques inconnues en se basant sur des modèles prédictifs.

Pour les fournisseurs à grande échelle, la décision de choisir une méthode de détection d'attaques DDoS dépendra de leur capacité à gérer des volumes massifs de trafic, à garantir la disponibilité des services et à maintenir le contrôle qu'ils souhaitent sur leur infrastructure de sécurité. Dans de nombreux cas, ces fournisseurs opteront pour des solutions personnalisées qui combinent plusieurs méthodes pour une protection efficace.

Les solutions de détection d'attaques DDoS volumétriques est un domaine en constante évolution en raison de la sophistication croissante des attaques DDoS. Les chercheurs et les professionnels de la sécurité travaillent constamment à l'amélioration des techniques de détection pour contrer ces attaques. La suite de cette section est un aperçu des solutions de détection d'attaques DDoS volumétriques que j'ai étudié pendant ma thèse.

Chapitre 2

Problématique

Sommaire

Dans ce chapitre, j'explore les enjeux et les défis spécifiques qui sous-tendent notre recherche sur la détection d'attaques DDoS volumétriques. Il est question de mettre en évidence les questions clefs que nous cherchons à résoudre dans le cadre ce manuscrit, en clarifiant pourquoi cette problématique est d'une importance curcial dans le domaine de la sécurité réseau des fournisseurs.

2.1 Détection DDoS dans chez fournisseur

La question qui a motivé ces travaux de recherche est la suivante : « *Comment détecter les attaques DDoS au sein de l'infrastructure d'un fournisseur de services cloud de grande envergure ?* »

2.1.1 Impacts

Comme nous l'avons exploré dans 1.3, une attaque DDoS est une forme de cyberattaque visant à rendre nos ressources réseau indisponibles en les submergeant de trafic illégitime. L'objectif est de perturber nos services pour les rendre inopérants. Une attaque DDoS se traduit par un assaillant submergeant notre infrastructure de trafic Internet non désiré, perturbant ainsi le flux de trafic légitime vers sa destination.

Pour mieux illustrer cette idée, nous pouvons la comparer à un embouteillage soudain provoqué par des centaines de fausses demandes de covoiturage. Ces demandes semblent authentiques pour nos services de covoiturage, qui envoient donc des conducteurs pour prendre en charge les passagers, engorgeant ainsi nos routes. Le trafic légitime est entravé, et nos utilisateurs ne peuvent pas atteindre leurs destinations.

Lorsque nous évoquons les attaques DDoS volumétriques, nous faisons référence à des attaques basées sur un volume de trafic pouvant atteindre plusieurs térabits par seconde, conçues pour saturer la bande passante de notre infrastructure. Ces attaques ont des conséquences significatives tant pour nous, en tant que fournisseur de services *cloud*, que pour nos clients, car elles affectent l'ensemble de notre écosystème.

Enfin, lorsque nous parlons de nos infrastructures *cloud*, nous faisons référence à nos propres infrastructures matérielles, chez OVHcloud par exemple. Ces infrastructures englobent tous les composants nécessaires au *cloud computing*, y compris les ressources réseau, le stockage, ainsi que les ressources matérielles abstraites pour la virtualisation. En termes simples, notre infrastructure *cloud* fournit tous les outils nécessaires pour créer et héberger des services et des applications au sein d'une même structure. Notre infrastructure est accessible via Internet, mais, comme nous l'avons détaillé dans la 1.1, elle se décline en plusieurs modèles de prestations. Pour fournir des services informatiques en tant que service, il est essentiel de comprendre que plusieurs composants doivent collaborer, notamment le réseau, le stockage, et la puissance de calcul. Tous ces éléments contribuent à la réussite de notre déploiement.

2.1.2 Répercussions

La détection des attaques DDoS volumétriques est importante pour nous en raison des répercussions significatives qu'elles peuvent engendrer. Ces attaques sont conçues

pour saturer la bande passante des services en ligne, les rendant ainsi indisponibles pour nos utilisateurs légitimes. La capacité à détecter ces attaques nous permet de déployer rapidement des mesures d'atténuation, réduisant ainsi l'impact sur la disponibilité de nos services. Nos clients s'attendent à ce que leurs services restent continuellement accessibles, et les attaques DDoS réussies ont le potentiel de miner leur confiance en notre fiabilité, affectant ainsi notre réputation.

L'identification rapide et l'intervention efficace jouent un rôle essentiel dans le maintien de la confiance de nos clients, tout en allégeant la charge de travail opérationnelle. Sans une détection adéquate, nos équipes de sécurité pourraient se retrouver submergées, entraînant des coûts opérationnels supplémentaires.

Les retombées de ces attaques sont multiples. Les attaques DDoS volumétriques provoquent une congestion massive de la bande passante, perturbant l'accès de nos utilisateurs légitimes à nos services en ligne. Ces périodes d'indisponibilité ont des conséquences préjudiciables pour nos clients qui dépendent de nos services pour leurs opérations. Les interruptions de service peuvent entraîner des pertes de revenus pour nos clients, notamment si leur activité repose sur nos services pour des transactions en ligne ou des opérations commerciales. En réponse à ces attaques, nous investissons dans des solutions de détection et d'atténuation des attaques DDoS, ce qui peut entraîner des coûts substantiels en termes de matériel, de logiciels et de personnel spécialisé. Nos clients, mécontents des perturbations, peuvent être tentés de rechercher des alternatives qui peuvent sembler plus fiables, augmentant ainsi le risque de perte de clientèle. Les attaques DDoS réussies ont également le pouvoir de ternir notre réputation, car nos clients peuvent perdre confiance en notre capacité à protéger leurs services contre de telles attaques. De plus, il n'est pas rare que les attaquants déploient une attaque DDoS de grande envergure à notre encontre dans le but de surcharger nos systèmes et d'attirer l'attention de nos équipes de sécurité. Pendant que nos équipes de sécurité se mobilisent pour contrer l'attaque DDoS, les attaquants peuvent simultanément mener des attaques plus insidieuses, comme l'infiltration de réseaux, l'exploitation de vulnérabilités, la compromission de comptes ou le vol de données.

2.1.3 Défis

Malgré des décennies d'études consacrées aux attaques DDoS dans la littérature scientifique, on peut constater que le problème persiste sans résolution. Un examen empirique révèle que les solutions recommandées dans l'état de l'art en matière de détection sont largement sous-utilisées, du moins dans le contexte d'OVHcloud. En explorant la littérature spécialisée, cette disparité ne résulte pas d'un retard de la part des acteurs industriels, mais plutôt de l'inadéquation des solutions proposées par rapport aux besoins des fournisseurs de services de grande envergure.

En général, les solutions que j'ai étudiées sont conçues pour des infrastructures de petite envergure, où les impacts des attaques sont relativement faciles à identifier. Par exemple, ces solutions sont efficaces lorsque les attaques sont de plusieurs ordres de grandeur plus importantes que le débit moyen du trafic nominal, ou lorsqu'elles altèrent radicalement les caractéristiques du trafic observé, telles qu'une surabondance du protocole utilisé pour mener l'attaque. Cependant, ces situations ne se produisent pas dans des infrastructures de la taille d'OVHcloud. Même les attaques dépassant plusieurs téraoctets de trafic par seconde ne représentent qu'une fraction minime du trafic total. C'est pourquoi, dans la suite de ce travail, nous proposons d'explorer trois questions clés visant à répondre à la problématique centrale précédemment évoquée.

Dans ce contexte, plusieurs questions se posent naturellement. Tout d'abord, il est impératif de comprendre : « *Pourquoi l'état de l'art académique diffère-t-il de l'état de l'art industriel ?* » Cette analyse permettra d'identifier les divergences entre les avancées théoriques en laboratoire et les défis pratiques rencontrés par les fournisseurs de services. Une fois cette question abordée, il faut se pencher sur la question suivante : « *Comment rendre accessibles des jeux de données 'réalistes' en matière de cloud ?* » Ces jeux de données doivent refléter les scénarios rencontrés par les acteurs industriels. Enfin, nous serons en mesure de répondre à la question : « *Comment détecter efficacement les attaques DDoS volumétriques dans un contexte de grande envergure ?* » Cette démarche éclairera les pistes à suivre pour combler le fossé entre la recherche académique et les besoins opérationnels de l'industrie.

2.2 État de l'art académique et industriel

Dans cette section, je vais présenter le contexte et l'interêt d'étudier la première question présentée à la section précédente : « *Comment détecter efficacement les attaques DDoS volumétriques dans un contexte de grande envergure ?* »

2.2.1 Défis

Lorsque nous examinons les jeux de données fréquemment utilisés par les chercheurs pour élaborer des solutions de détection des attaques DDoS volumétriques, il devient évident que la plupart de ces jeux de données sont désuets, notamment en ce qui concerne un fournisseur comme OVHcloud, étant donné qu'ils ont été collectés avant l'avènement de l'ère du *cloud computing*. De plus, il est essentiel de noter que la plupart de ces jeux de données ne correspondent pas aux caractéristiques dimensionnelles de notre infrastructure. Cette disparité provient du fait que les infrastructures *cloud* sont généralement gérées par un nombre limité d'acteurs qui sont réticents à partager des informations qu'ils considèrent comme relevant du secret industriel.

De plus, les réglementations actuelles, bien qu'essentielles pour la protection de nos données personnelles, peuvent constituer un obstacle significatif à la diffusion de jeux de données réalistes. Les entreprises sont soumises à des obligations strictes en matière de confidentialité des données, ce qui limite leur capacité à partager des informations sur les attaques et les vulnérabilités rencontrées, même à des fins de recherche.

2.2.2 Réproductibilité

Ce sujet est important à plusieurs égards. Tout d'abord, il joue un rôle clé dans l'assurance de la reproductibilité des recherches entreprises et des solutions qui en résultent. La reproductibilité est un pilier essentiel de la méthodologie scientifique, car elle permet à d'autres chercheurs de valider, de vérifier et de bâtir sur des travaux antérieurs. En garantissant que les données et les conditions expérimentales sont accessibles et bien documentées, la communauté de recherche peut s'assurer que les découvertes sont robustes et peuvent être confirmées par d'autres, contribuant ainsi à l'avancement des connaissances.

De plus, cette question est cruciale pour faciliter l'amélioration continue et le test des solutions développées. En fournissant des jeux de données réalistes et des contextes d'expérimentation représentatifs, la recherche en sécurité informatique peut s'aligner sur les conditions réelles auxquelles les technologies de détection des attaques DDoS volumétriques seront confrontées. Cela permet de concevoir, d'optimiser et de valider des solutions plus efficaces, qui sont mieux adaptées aux besoins des fournisseurs. En somme, un accès facile et transparent à des données et des environnements de test réalistes est un catalyseur essentiel pour faire progresser la sécurité en ligne et assurer la fiabilité des systèmes face aux menaces numériques.

2.2.3 Disparité des jeux de données

Afin de répondre à la question qui se pose ici, nous devons explorer plusieurs sous-questions, chacune contribuant de manière spécifique à la résolution globale du problème.

Tout d'abord, nous devons nous pencher sur la question de savoir : « *Quels sont les jeux de données les plus couramment utilisés ?* » Cette interrogation vise à acquérir une compréhension approfondie des sources d'information qui sont fréquemment exploitées au sein de la communauté de recherche. Identifier les jeux de données couramment utilisés nous permettra de cerner la base de données sur laquelle repose une grande partie de la recherche en détection des attaques DDoS, tout en mettant en lumière les éventuelles lacunes.

Deuxièmement, il est essentiel d'explorer la question suivante qui découle de la première : « *Quelles sont les métriques les plus utilisées pour caractériser les attaques DDoS ?* » La définition et l'utilisation de métriques standardisées sont cruciales pour caractériser les attaques et comparer les solutions de détection. Ces métriques fourniront les outils nécessaires pour analyser le trafic dans les jeux de données identifiés, ainsi que pour évaluer le trafic observé au sein de l'infrastructure d'OVHcloud. Cette analyse mettra en évidence les disparités entre les jeux de données et le trafic d'un fournisseur de services de grande envergure.

Cela nous amène à la question suivante : « *Comment capturer le trafic au sein d'une infrastructure telle qu'OVHcloud ?* » Les infrastructures de cette ampleur génèrent un volume considérable de trafic, souvent exprimé en téraoctets par seconde. Il faut trouver des méthodes de capture qui permettent de recueillir les données nécessaires sans perturber le bon fonctionnement des opérations.

Un détail est que, compte tenu de la quantité massive de trafic généré au sein d'un fournisseur de services, le trafic est généralement échantillonné. Il est donc essentiel de déterminer : « *Quel est l'impact de l'échantillonnage sur les métriques que nous avons sélectionnées ?* » Cette question garantira que nous pouvons comparer de manière valide les jeux de données issus de la littérature avec le trafic observé au sein de notre infrastructure, malgré l'échantillonnage.

Finalement, la dernière question à résoudre est : « *Comment comparer les résultats issus des deux mondes ?* » Cette interrogation nous permettra de tirer des conclusions de nos observations et de proposer des solutions pour combler les divergences constatées. En réunissant les éléments recueillis à travers ces sous-questions, nous serons en mesure de jeter un éclairage précis sur la disparité entre les jeux de données académiques et la réalité opérationnelle des fournisseurs, ouvrant la voie à des solutions plus adaptées et efficaces dans la détection des attaques DDoS volumétriques.

2.3 Disponibilité des jeux de données

Ici, je présente le contexte ainsi que les implications liées à la question : « *Comment rendre disponible des jeux de données "réalistes" de fournisseurs de services cloud ?* »

2.3.1 Défis de la disponibilité

La question de savoir comment rendre disponibles des jeux de données « réalistes » de fournisseurs est un enjeu complexe et essentiel pour la recherche en détection d'attaques DDoS volumétriques. Tout d'abord, la notion de jeux de données « réalistes ». Il s'agit de données qui capturent fidèlement la variété des scénarios et des caractéristiques du trafic et des attaques dans un environnement de fournisseur. Ces données

doivent refléter la complexité et la diversité du trafic, des charges de travail et des attaques observés dans la réalité opérationnelle, ce qui représente un défi en soi.

Ensuite, les fournisseurs jouent un rôle central dans cette question. Les jeux de données réalistes doivent provenir de ces fournisseurs pour être représentatifs des conditions auxquelles ils sont confrontés. Cependant, les fournisseurs de services *cloud* sont souvent réticents à partager leurs données en raison de préoccupations liées à la confidentialité, à la sécurité et à la protection des informations sensibles de leurs clients. Il est donc nécessaire de trouver un équilibre entre l'accès aux données et la préservation de la confidentialité.

La disponibilité des jeux de données est également un point clé. Il s'agit de mettre à disposition ces données de manière à ce qu'elles soient accessibles aux chercheurs et aux acteurs de l'industrie. La disponibilité peut être publique, restreinte ou contrôlée, en fonction des accords et des politiques en place. Cependant, il est impératif de clarifier les modalités d'accès pour que les chercheurs puissent utiliser ces données de manière éthique et en conformité avec les réglementations sur la confidentialité des données.

En fin de compte, cette question nécessite des efforts conjoints des chercheurs, des fournisseurs, des organismes de réglementation et des acteurs de l'industrie pour élaborer des protocoles et des normes de partage de données qui préservent la sécurité et la confidentialité, tout en favorisant la recherche en détection d'attaques DDoS volumétriques. La création de jeux de données « réalistes » accessibles et pertinents est essentielle pour améliorer la sécurité des services en ligne face à ces menaces persistantes.

2.3.2 Importance des jeux de données réalistes

La disponibilité de jeux de données réalistes provenant de fournisseurs revêt une importance dans le domaine de la recherche scientifique en détection d'attaques DDoS volumétriques. Ces jeux de données reflètent les conditions réelles auxquelles sont confrontées ces infrastructures, garantissant ainsi la représentativité de la réalité opérationnelle. Ils contiennent une variété d'attaques DDoS ainsi que les caractéristiques spécifiques du trafic, des charges de travail et des comportements des utilisateurs. En utilisant ces données, les chercheurs sont en mesure de travailler avec des scénarios de test authentiques, ce qui améliore la pertinence de leurs résultats de recherche.

De plus, ces jeux de données réalistes facilitent l'évaluation de l'efficacité des solutions de détection. En simulant des situations du monde réel, les chercheurs peuvent déterminer la performance, la sensibilité, la spécificité et les taux de détection des solutions de manière plus fiable. Ainsi, ils peuvent identifier les technologies les plus appropriées pour la protection des infrastructures contre les attaques DDoS.

Les jeux de données réalistes sont également essentiels pour valider les modèles de

recherche. Ils permettent de vérifier la robustesse des approches proposées en utilisant des données authentiques, tout en assurant que les hypothèses sous-jacentes aux modèles sont en accord avec la réalité. Les chercheurs peuvent ainsi affiner leurs modèles de détection et les adapter à des conditions opérationnelles spécifiques.

De plus, ces jeux de données favorisent le développement de solutions adaptées aux fournisseurs de services *cloud*. En considérant les particularités de ces environnements, les chercheurs peuvent concevoir des techniques de détection plus efficaces, tenant compte des volumes massifs de trafic, de la diversité des services en ligne et des méthodes d'atténuation appropriées.

Enfin, la disponibilité de jeux de données réalistes encourage la collaboration entre la communauté de recherche et l'industrie. Les fournisseurs sont plus enclins à partager des données lorsque cela contribue à renforcer la sécurité de leurs services. Cette collaboration permet aux chercheurs de bénéficier d'informations en temps réel, de mieux comprendre les tendances actuelles en matière d'attaques DDoS, et de mettre au point des solutions en phase avec les besoins de l'industrie.

2.3.3 Défis de l'accès

Pour répondre à cette question, il est nécessaire de se pencher sur une série de préoccupations clés. Tout d'abord, la question de : « *Comment donner accès à des données aussi sensibles ?* » se pose inévitablement. En effet, les données requises pour créer des jeux de données réalistes sont souvent constituées d'informations sensibles qui sont soumises à diverses réglementations sur la confidentialité des données. La protection de la vie privée et la conformité légale sont des facteurs cruciaux à prendre en compte lors de l'accès et de la manipulation de ces données. Il est impératif de mettre en place des mécanismes de protection des données et d'anonymisation pour garantir que les informations sensibles ne sont pas exposées.

Ensuite, il faut se demander : « *Comment s'assurer que les données sont au moins statistiquement proches de ce que l'on observe sur l'infrastructure ?* » En effet, la représentativité des données est une question importante. Si les données ne reflètent pas avec précision la réalité opérationnelle d'un fournisseur, comme OVHcloud par exemple, elles risquent de s'avérer contre-productives. Il est donc nécessaire de s'appuyer sur des travaux de modélisation qui émanent de la question précédente, à savoir : « *Pourquoi l'état de l'art académique diffère de l'état de l'art industriel ?* » L'analyse des divergences entre les environnements académiques et industriels offre des pistes pour ajuster les jeux de données en fonction des caractéristiques spécifiques des fournisseurs.

Finalement, il faut mettre en place des outils techniques permettant de reproduire les données observées, répondant ainsi à la question : « *Est-il possible d'utiliser un outil déjà connu de l'état de l'art pour fournir ce jeu de données, ou faut-il en proposer*

un nouveau ? » La création d'outils de génération de données réalistes est essentielle pour produire des ensembles de données qui reflètent fidèlement les conditions d'un environnement de fournisseur. Cela peut nécessiter des approches novatrices pour la modélisation, la génération et l'évaluation des données.

En résumé, la mise à disposition de jeux de données réalistes de fournisseurs de services *cloud* soulève des questions complexes liées à la confidentialité des données, à la représentativité, à la modélisation et à la création d'outils techniques. En répondant de manière réfléchie à ces interrogations, il est possible de garantir que les données mises à disposition sont à la fois précieuses pour la recherche en détection d'attaques DDoS volumétriques et conformes aux réglementations sur la confidentialité des données.

2.4 Détection des DDoS volumétrique chez un fournisseur

Dans cette section, je vais discuter du contexte et de l'importance de la dernières des trois sous questions : « *Comment détecter efficacement les attaques DDoS volumétriques dans un contexte de grande envergure ?* »

2.4.1 Défis

Les attaques DDoS, sont des attaques informatiques conçues pour rendre un service en ligne, un site web ou une application indisponibles comme nous l'avons vu dans 1.3.

La détection dans le domaine de la sécurité informatique consiste à repérer précocement toute menace ou activité malveillante. Dans le cas des attaques DDoS, la détection implique une surveillance constante du trafic réseau, à la recherche de signaux ou de comportements qui pourraient indiquer une attaque en cours. Cela inclut l'analyse des motifs de trafic, la détection d'anomalies dans le comportement des utilisateurs, et même la surveillance des performances des serveurs pour détecter des signes d'attaque potentielle.

L'efficacité de la détection revêt une grande importance dans ce contexte. Elle désigne la capacité à identifier rapidement et avec précision les attaques DDoS tout en minimisant les faux positifs, c'est-à-dire les alertes incorrectes signalant une activité normale comme une attaque. L'efficacité requiert un équilibre délicat entre la détection précise des attaques réelles et la réduction des perturbations inutiles pour les utilisateurs légitimes, car des alertes excessives peuvent également entraîner des problèmes.

Les attaques DDoS volumétriques se distinguent par l'envoi massif de trafic vers la cible, souvent dépassant largement la capacité de cette dernière à le gérer. Ces attaques visent à submerger les ressources du système en saturant la bande passante ou en engorgeant les capacités de traitement. Les attaques volumétriques sont souvent les plus dévastatrices, étant capables de causer des temps d'arrêt significatifs en seulement quelques minutes.

Enfin, le contexte de grande envergure renvoie à une infrastructure informatique complexe et étendue, comprenant de nombreux serveurs, réseaux, services et utilisateurs, fonctionnant à une grande échelle comme celle d'OVHcloud. Dans un environnement aussi complexe, la détection des attaques DDoS peut s'avérer particulièrement difficile en raison de la diversité des points d'entrée possibles et de la variété du trafic. Les systèmes de détection doivent être spécifiquement adaptés pour gérer cette complexité et être véritablement efficaces.

2.4.2 La détection d'attaques

Pour les fournisseurs, la détection efficace des attaques DDoS volumétriques est d'une importance critique. Ces entreprises offrent une gamme diversifiée de services d'hébergement et de stockage de données pour un large éventail de clients, allant des petites entreprises aux grandes organisations. Dans ce contexte, la disponibilité et la sécurité de ces services revêtent une importance capitale.

Les attaques DDoS, avec leur capacité à inonder les serveurs de trafic malveillant, représentent une menace sérieuse. Une attaque DDoS réussie peut causer des temps d'arrêt, perturber les opérations des clients, et nuire à la réputation du fournisseur. Par conséquent, le développement de mécanismes de détection efficaces pour repérer ces attaques dès leur apparition est impératif.

La première raison pour laquelle les fournisseurs optent souvent pour le développement de solutions de détection en interne réside dans leur responsabilité envers les clients. En tant que gardiens des données et des applications de leurs clients, ces fournisseurs ont une obligation envers la continuité des services. La détection précoce et la réponse rapide aux attaques DDoS garantissent que les clients peuvent continuer à accéder à leurs données et à leurs applications sans interruption majeure.

De plus, les attaques DDoS peuvent non seulement entraîner des temps d'arrêt, mais également impacter négativement les performances globales de l'infrastructure. Le développement de solutions de détection en interne permet aux fournisseurs de mieux gérer la performance de leurs services, en limitant les effets des attaques sur l'ensemble de la plateforme.

Les avantages financiers sont également un argument de poids en faveur du développement interne de solutions de détection. Les solutions tierces peuvent s'avérer coûteuses à long terme, tandis que le développement en interne permet de contrôler les coûts

tout en répondant précisément aux besoins de l'entreprise. De plus, la personnalisation et l'adaptabilité des solutions internes permettent de mieux répondre aux besoins spécifiques de l'entreprise et d'évoluer avec les menaces en constante évolution.

La confidentialité des données et le contrôle sur les mécanismes de détection sont également des facteurs essentiels pour nous, fournisseur. Les données sensibles peuvent être impliquées dans le processus de détection, et le développement en interne garantit un contrôle total sur l'accès à ces données, renforçant la sécurité et la confidentialité.

Enfin, la capacité à détecter et à réagir efficacement aux attaques DDoS peut constituer un avantage concurrentiel significatif pour les fournisseurs. Cela renforce leur réputation et leur crédibilité, attirant de nouveaux clients et fidélisant les clients existants, dans un marché hautement concurrentiel.

En conclusion, la détection efficace des attaques DDoS dans un contexte de grande envergure est essentielle pour les fournisseurs, qui sont les gardiens des données et des services de leurs clients. Le développement de solutions de détection en interne offre un meilleur contrôle, une personnalisation adaptée aux besoins spécifiques et une meilleure rentabilité, renforçant ainsi leur position sur le marché compétitif des services *cloud*.

2.4.3 Validation

La validation d'un système de détection des attaques DDoS dans un environnement de grande envergure est une étape critique pour garantir la robustesse de la sécurité en ligne. Cette méthodologie repose sur l'exploration de questions pour s'assurer que le système est capable de faire face aux défis spécifiques rencontrés par les fournisseurs.

La première question porte sur : « *Comment gérer le volume de trafic important d'un fournisseur de services cloud ?* » La validation doit tenir compte de la capacité du système à gérer de manière efficace un trafic intensif. Également, la minimisation des faux positifs est une autre préoccupation majeure. Il faut se demander : « *Comment minimiser les faux positifs ?* » Un système de détection doit être précis pour éviter de déclencher des alertes injustifiées. Mais aussi, un autre aspect à prendre en compte sont les pics soudains de trafic légitime, les *flash-crowds*, « *Comment gérer les Flash-Crowds ?* », nécessitant également une attention particulière. Enfin, la détection des attaques DDoS volumétriques de faible débit est un défi dans notre contexte : « *Comment détecter même les DDoS Volumétriques de plus faible débit ?* » pour assurer de détecter le plus d'attaques possible.

2.5 Conclusion

La détection efficace des attaques DDoS dans un contexte de grande envergure, tel que celui des fournisseurs de services *cloud*, est importante pour garantir la disponibilité et la sécurité des services en ligne. Cette discussion a mis en lumière les défis spécifiques auxquels sont confrontés, ainsi que les raisons pour lesquelles ils optent souvent pour le développement de solutions de détection en interne. La validation de ces systèmes de détection est essentielle pour s'assurer de leur robustesse dans un environnement complexe.

Plusieurs questions clés ont été abordées. Il est apparu que la gestion du volume de trafic massif, la minimisation des faux positifs, la gestion des *flash-crowds* et la détection des attaques de faible débit sont autant de défis auxquels les systèmes de détection doivent faire face. La résolution de ces questions est indispensable pour garantir la fiabilité des mécanismes de détection des attaques DDoS.

Il a également été mis en évidence l'importance de l'accès à des jeux de données réalistes pour la recherche en détection des attaques DDoS volumétriques. Ces jeux de données reflètent les conditions opérationnelles des fournisseurs et permettent aux chercheurs de développer, tester et valider des solutions de détection plus efficaces.

Finalement, la discussion a souligné que la sécurité en ligne est une préoccupation partagée par les fournisseurs, les chercheurs et l'industrie en général. Le développement de solutions de détection efficaces et la collaboration entre ces acteurs sont essentiels pour faire face à la menace persistante des attaques DDoS volumétriques.

En résumé, la détection des attaques DDoS dans un contexte de grande envergure est un défi de taille, mais des réponses appropriées aux questions clés, une validation rigoureuse et une collaboration continue peuvent permettre de renforcer la sécurité des services en ligne face à ces menaces toujours présentes.

Chapitre 3

Modélisation statistique du trafic pour la détection d'attaques DDoS

Sommaire

L'état actuel de la recherche met en évidence un problème crucial : le manque de données représentatives concernant les fournisseurs de services *cloud*. Cette carence de données se traduit dans la littérature scientifique par des solutions de détection d'attaques DDoS volumétriques qui sont souvent jugées irréalistes par les ingénieurs travaillant au sein d'entreprises *cloud*, notamment ceux avec lesquels j'ai eu l'occasion de discuter au sein d'OVHcloud.

Nous avons entrepris une démarche visant à caractériser le trafic circulant au sein du cœur du réseau (*backbone*) en utilisant des statistiques couramment référencées dans la littérature. Cette caractérisation nous permet de mieux comprendre à quoi ressemble le trafic normal ainsi que les attaques auxquelles les fournisseurs de services *cloud* sont confrontés au quotidien. De plus, cette étude nous permet d'appréhender plus précisément les défis auxquels sont confrontés ces fournisseurs, tels que la distinction entre le trafic utilisateur ou client et le trafic interne ou d'administration, la gestion de la croissance rapide du trafic, la distribution géographique des points

d'accès, ainsi que la complexité de l'écosystème réseau avec ses technologies d'échantillonnage, d'encapsulation et d'équilibrage de charge, qui rendent la modélisation du trafic particulièrement ardue.

Ainsi, dans ce chapitre, nous allons explorer différents jeux de données issues de la littérature, ainsi que les métriques fréquemment mentionnées, afin de caractériser le trafic. Notre objectif est de produire une modélisation du trafic d'OVHcloud, ce qui nous permettra de mieux comprendre les divergences entre les jeux de données disponibles et les réalités auxquelles un fournisseur de services *cloud* comme OVHcloud est confronté au quotidien. Cette approche contribuera à combler le fossé entre la recherche académique et les besoins pratiques de l'industrie des services *cloud* en matière détection d'attaques DDoS.

3.1 Sélection des jeux de données

Les chercheurs spécialisés dans le domaine des systèmes de détection d'intrusion (IDS), en se concentrant particulièrement sur la détection des attaques DDoS de grande ampleur, ont généralement recours à des ensembles de données provenant de captures de trafic réseau réelles ou de simulations partielles ou complètes d'environnements [BK16]. Dans le but de proposer une modélisation du trafic d'OVHcloud afin de soutenir le développement de solutions axées sur les fournisseurs de services cloud, ma première étape consiste à répertorier les différents ensembles de données les plus couramment rencontrés dans la littérature scientifique. Ensuite, je vais m'appuyer sur ces ensembles de données identifiés pour créer une modélisation du trafic qui pourrait servir de base à la création d'un nouvel ensemble de données dédié à cette fin.

Pour recenser les jeux de données couramment utilisés dans la littérature, nous avons examiné différentes études, telles que celles de Ring et al. [Rin+19], Camargo et al. [Cam+18] et Damasevicius et al [Dam+20]. En utilisant ces études comme références, nous avons établi une liste de jeux de données, compilée dans le tableau 3.1 que nous considérons représentatifs de ceux utilisés dans la littérature.

Cependant, comme nous pouvons l'observer dans le tableau 3.1, la plupart des jeux de données sont trop anciens ; ils datent d'avant l'utilisation du *cloud computing* pour être utilisés dans des études récentes montrant les effets des attaques DDoS sur les fournisseurs. Comme mentionné dans l'état de l'art, les attaques DDoS ont évolué au fil du temps, tout comme les usages légitimes. Par exemple, les jeux de données *DARPA* et *KDD CUP 99* remontent à une période antérieure aux infrastructures massivement distribuées et des services de *DDoS-as-a-Service*, ainsi que des applications qui servent des ressources à des millions d'utilisateurs à travers le monde.

Il convient également de noter que de nombreux jeux de données ne sont pas

Jeux de données	Année	Taille	Durée	Format	Public	Étudié	Trafic d'attaque	Trafic légitime
KDD CUP 99 [Tav+09]	1998	1GB	5 semaines	Connexions	Oui	Non	Oui	Oui
PU-IDS [SKS15]	1998	1GB	N/S	N/S	Oui	Non	Oui	Oui
DARPA [TSB08]	1999	10GB	5 semaines	Paquets	Oui	Non	Oui	Oui
Kyoto 2006+ [Son+11]	2006	100GB	9 ans	Bro IDS sessions	Oui	Non	Oui	Oui
NSL-KDD [Pro18]	2009	1GB	5 semaines	Connexions	Oui	Non	Oui	Oui
SSENET-2011 [VHS11]	2011	N/S	4 heures	N/S	Non	Non	Oui	Oui
TUIDS [Gog+12]	2012	N/S	21 jours	Paquets et Flux	Non	Non	Oui	Oui
ISCX 2012 [KES21]	2012	85GB	7 jours	Paquets	Oui	Oui	Oui	Oui
Booters [San+15b]	2013	250GB	2 jours	Paquets	Oui	Non	Oui	Non
SSENET-2014 [BS14]	2014	N/S	4 heures	N/S	Non	Non	Oui	Oui
Santa [Whe+14]	2014	N/S	N/S	N/S	Non	Non	Oui	Oui
UNSW-NB15 [MS15]	2015	150GB	31 heures	Paquets	Oui	Oui	Oui	Oui
NDSec-1 [Cat+21]	2016	2GB	N/S	Paquets	Oui	Non	Oui	Non
DDoS 2016 [Alk+16]	2016	1GB	N/S	Résumé de paquets	Oui	Non	Oui	Oui
URG'16 [Mac+18]	2016	14GB	4 mois	Netflow	Oui	Oui	Oui	Oui
CIC IDS 2017 [SLG18]	2017	50GB	5 jours	Paquets	Oui	Non	Oui	Oui
CIC DoS [Jaz+17]	2017	5GB	1 jour	Paquets	Oui	Oui	Oui	Oui
CSE-CIC-DDoS [SLG18]	2018	500GB	2 jours	Paquets	Oui	Oui	Oui	Oui
DDoS 2019 [Sha+19]	2019	150GB	N/S	Paquets	Oui	Oui	Oui	Oui
LITNET-2020 [Dam+20]	2020	1GB	10 mois	Netflow	Oui	Oui	Oui	Oui

TABLE 3.1 – Résumé des jeux de données

disponibles publiquement, ce qui pose un problème pour la reproductibilité et le développement d'expériences de détection des attaques DDoS volumétriques.

Enfin, il est important de souligner le fossé entre la volumétrie des jeux de données couramment utilisés et celle observée sur l'infrastructure d'OVHcloud. Par exemple, un dataset au format *PCAP* d'une durée de plusieurs heures, comprenant les en-têtes et les charges utiles des paquets, ne dépasse généralement pas quelques dizaines de gigaoctets, alors qu'au moment de la rédaction de ce mémoire, plusieurs dizaines de térabits de trafic sont observés par secondes sur la backbone d'OVHcloud, soit trois ordres de grandeurs de différence. Des chiffres qui mettent en évidence l'ampleur des défis liés à la collecte et à la modélisation des données dans les infrastructures des fournisseurs de sevices *cloud*.

Pour la suite de notre étude, nous avons sélectionné les jeux de données suivants : *ISCX 2012*, *UNSW-2015*, *URG'16*, *CIC IDS 2017*, *CSE-CIC-DDoS* et *LITNET-2020*. Dans un premier temps, nous allons comparer ces jeux de données en utilisant l'évolution du nombre maximal d'adresses IP sources par destination au fils du temps, le nombre de paquets total par minutes observés, la répartion des protocoles TCP, UDP, $ICMP$ dans le temps ainsi que l'évolution des entropies des ports source et destination, à la fois entre eux et par rapport aux valeurs observées sur le trafic de production d'OVHcloud.

ISCX 2012 souvent utilisé dans la recherche en cybersécurité, est une ressource précieuse pour l'analyse et l'évaluation des menaces en ligne. Il a été créé pour capturer des données de trafic réseau réel dans un environnement de laboratoire contrôlé. L'ensemble de données comprend une variété de scénarios, notamment des activités normales et des attaques, ce qui en fait un outil puissant pour développer et tester des techniques de détection d'anomalies et d'attaque. $ISCX2012$ couvre une gamme diversifiée d'attaques, telles que DDoS, les attaques par injection SQL, les attaques par force brute, et bien d'autres.

URG'16 est construit avec du trafic réel et des attaques récentes. Ces données proviennent de plusieurs collecteurs *Netflow* stratégiquement positionnés dans le réseau d'un FAI espagnol. Il se compose de deux ensembles de données distincts qui ont été préalablement divisés par semaine

CIC IDS 2017 L'ensemble de données contient des activités normales ainsi que les attaques courantes les plus récentes, ce qui ressemble aux données réelles du monde réel. Cet ensemble de données contient un large éventail de scénarios de trafic réseau, y compris des activités normales et des attaques, ce qui en fait un outil précieux pour développer et tester des algorithmes de détection d'anomalies et d'attaques.

CSE-CIC-DDoS est une collection d'attaques DDoS synthétiques et réelles, en-

registrées dans un environnement de laboratoire. Il offre une variété de scénarios d'attaques DDoS, ce qui en fait un choix approprié pour évaluer la capacité de détection des systèmes IDS.

LITNET-2020 est un ensemble de données récent qui fournit des captures de trafic réseau réel issue du trafic d'une université, y compris des attaques DDoS volumétriques. Il est caractérisé par sa pertinence en termes de représentativité des attaques actuelles observées dans les infrastructures réseau.

Nous comparerons ces jeux de données en examinant des mesures telles que les volumes de trafic, les types d'attaques présentes, la diversité des sources et des cibles des attaques, ainsi que d'autres métriques pertinentes. De plus, nous les confronterons aux valeurs évoquées précédement que nous pouvons observer sur le trafic de production d'OVHcloud, qui représente un environnement réel et dynamique.

Cette comparaison nous permettra d'évaluer la pertinence et la représentativité des différents jeux de données sélectionnés pour nos expérimentations et de mieux comprendre leur adéquation à la réalité du trafic sur les infrastructures fournisseurs de sevices *cloud*, telles que celle d'OVHcloud.

3.2 Sélection des metriques

Dans le but d'identifier les points de divergence entre le trafic de production d'OVHcloud et les jeux de données disponibles dans la littérature, nous avons sélectionné des métriques utilisées dans l'état de l'art. Nous avons choisi les métriques suivantes :

Nombre d'adresses IP source par adresse IP destination Cette métrique permet de quantifier la diversité des sources et des cibles dans un trafic donné. Dans le cas des attaques DDoS, il est courant que de nombreux ordinateurs compromis (appartenant à un botnet) envoient des paquets vers une ou plusieurs cibles. Ainsi, une valeur élevée de cette métrique peut indiquer la présence d'une attaque DDoS où plusieurs adresses IP sources sont impliquées dans l'attaque contre une ou plusieurs adresses IP destinations. Cela permet d'évaluer la capacité des jeux de données à représenter cette diversité d'attaques et à détecter les schémas caractéristiques des attaques DDoS. En outre, cette métrique fournit également une indication de la complexité de l'infrastructure à partir de laquelle provient l'ensemble de données. Même lorsque le trafic est classé comme légitime, une observation importante réside dans le nombre élevé d'adresses IP source par rapport aux adresses IP de destination, suggérant ainsi que certains composants de l'infrastructure subissent une charge significative. Cette observation peut également indiquer que l'in-

frastructure examinée présente des similitudes avec celle d'un fournisseur de services *cloud* en termes d'importance.

Nombre total de paquets observés Cette métrique fournit une indication du volume global du trafic dans un dataset. Les attaques DDoS sont souvent caractérisées par un trafic anormalement élevé, générant un grand nombre de paquets qui peuvent submerger la capacité des infrastructures cibles. Par conséquent, en évaluant le nombre total de paquets dans les jeux de données, il est possible de comparer l'échelle du trafic des attaques DDoS simulées ou enregistrées avec celles observées dans le trafic de production d'OVHcloud ou d'autres fournisseurs de services cloud.

Distribution des protocoles La distribution des protocoles utilisés dans le trafic permet d'analyser les variations entre les différents jeux de données. Les attaques DDoS peuvent utiliser différents protocoles pour cibler les infrastructures. En évaluant la distribution des protocoles dans les jeux de données, on peut déterminer si ces protocoles sont représentés de manière adéquate et si les attaques DDoS volumétriques sont bien présentes. Cela permet de comparer la composition du trafic des jeux de données avec celle du trafic réel et de détecter d'éventuelles différences significatives. Comme nous nous intéressons aux attaques DDoS volumétriques qui utilisent les couches 3 et 4 du modèle OSI, nous étudions la distribution des protocoles TCP, UDP et $ICMP$. D'autres protocoles peuvent être utilisés pour conduire des attaques DDoS, mais nous avons décidé de les regrouper dans une unique catégorie « $OTHER$ », puisque ce sont les vecteurs d'attaques les moins représentés dans les jeux de données de la littérature mais aussi les procotoles majoritairement observé sur l'infrastructure d'OVHcloud. Pour chaque fenêtre d'agrégation – Une fenêtre d'agrégation est une période de temps définie utilisée pour regrouper et calculer des valeurs agrégées à partir de données temporelles. – d'une minute, nous calculons donc le ratio entre les protocoles cités. Pour obtenir la valeur de cette métrique, dans un premier temps, il faut calculer le nombre d'occurrences ainsi que le ratio des protocoles étudiés dans notre fenêtre d'une minute. Il y a donc un compteur d'occurrence associé à chaque protocole, tel que : C_{TCP} : nombre de paquets TCP, C_{UDP} : nombre de paquets UDP, C_{ICMP} : nombre de paquets ICMP, C_{OTHER} : nombre de paquets avec d'autres protocoles (IP). Ensuite, nous calculons le nombre total de paquets dans la fenêtre d'agrégation d'une minute, désigné par Tc :

$$Tc = C_{TCP} + C_{UDP} + C_{ICMP} + C_{OTHER} \tag{3.1}$$

Maintenant, nous pouvons calculer les différents ratios de protocole R_p,

où $p \in \{TCP, UDP, ICMP, OTHER\}$:

$$R_p = \frac{C_p}{Tc} \tag{3.2}$$

La mesure distribution des protocoles peut être représentée par un vecteur ou un tuple des ratios des protocoles :

$$ProtocolDistribution = (R_{TCP}, R_{UDP}, R_{ICMP}, R_{OTHER}) \tag{3.3}$$

Ou $R_{TCP} + R_{UDP} + R_{ICMP} + R_{OTHER} = 1$. Ces ratios représentent la fréquence relative de chaque type de protocole dans la fenêtre temporelle d'une minute, et leur somme est égale à 1 car ils représentent ensemble 100% des paquets dans cette fenêtre.

Entropie des ports source et destination L'entropie des ports source et destination mesure la diversité des ports utilisés dans le trafic. Les attaques DDoS peuvent cibler des ports spécifiques pour exploiter des vulnérabilités ou simplement pour surcharger les ressources d'un système. Une entropie élevée indique une grande variété de ports utilisés, ce qui peut être indicatif de la présence d'attaques DDoS sophistiquées qui tentent de masquer leur activité malveillante en utilisant différents ports. L'évaluation de l'entropie des ports source et destination permet donc de détecter des schémas anormaux de ports-canning ou d'autres activités malveillantes associées aux attaques DDoS. Nous avons calculé l'entropie des ports sources et destinations en utilisant l'entropie de Shannon :

$$H(P) = -\sum_i p_i \log p_i \tag{3.4}$$

Où $H(P)$ est l'entropie de la distribution des ports, P est l'ensemble des ports uniques et p_i est la probabilité qu'un paquet utilise le port i.

Pour calculer ces métriques, nous avons opté, de manière expérimentale, pour une agrégation des informations par fenêtre d'une minute. Cette durée a été choisie comme un compromis entre le volume de données généralement disponible dans les jeux de données et le volume moyen de données transitant sur l'infrastructure d'OVHcloud. Ainsi, nous avons évalué ces métriques pour chaque dataset retenu dans la section précédente, puis nous les avons également calculées sur une capture du trafic de production d'OVHcloud.

L'évaluation de ces métriques nous permettra de mettre en évidence les différences entre les jeux de données existants et le trafic de production d'OVHcloud, fournissant ainsi un aperçu précieux de la pertinence et de la représentativité des jeux de données sélectionnés pour l'étude des attaques DDoS.

En utilisant ces métriques, il est possible de comparer les caractéristiques des attaques DDoS dans les jeux de données avec celles observées dans le trafic de production d'OVHcloud ou d'autres infrastructures réelles. Cela permet d'évaluer la pertinence et la représentativité des jeux de données pour l'étude des attaques DDoS, ainsi que d'identifier d'éventuelles différences significatives entre les jeux de données et le trafic réel.

3.3 Mise en oeuvre de la collecte

Afin de modéliser le flux de trafic régulier d'OVHcloud, une phase initiale de collecte a été entreprise.

Cependant, en raison du volume considérable de plusieurs térabits par seconde de trafic reçu par l'infrastructure, il était impossible de capturer exhaustivement chaque paquet de données (1 :1). Par conséquent, le système de collecte actuellement en production a été utilisé, basé sur la technologie *NetFlow* au niveau du cœur du réseau.

NetFlow permet la collecte et l'analyse d'informations clés sur le trafic réseau, telles que la quantité de données transférées, les adresses IP source et destination, les ports utilisés, les protocoles, les durées des sessions, etc. Bien que les NetFlows soient une représentation résumée des échanges réseau entre l'*Autonomous System*[1] d'OVHcloud et Internet, ils génèrent néanmoins une grande quantité de données. En effet, plus de 500 Go de *NetFlows* ont été capturés au cours de la période de collecte de 24 heures.

La collecte parallèle entreprise pour cette expérience devait être capable de traiter les flux en quasi-temps réel, sans introduire de latence dans le système de collecte en production. Il est à souligner que ce système de collecte n'a pas été spécifiquement conçu pour cette expérience, mais il est déjà utilisé par le système de détection d'attaques DDoS en production ainsi que par les équipes pour diverses opérations.

Pour le traitement des *NetFlows* reçus depuis le cluster *Kafka*[2], un programme *Rust*[3] a été développé. Ce programme se connecte au *topic*[4] dédié aux *NetFlows* émis par les équipements réseau de la backbone et extrait les informations pertinentes, telles que l'adresse IP source, l'adresse IP destination, le port source, le port

1. Un AS est un groupe de réseaux informatiques gérés par une même entité et qui utilise un protocole de routage commun sur Internet

2. Kafka est un logiciel de traitement de flux de données en temps réel, développé par la Fondation Apache, utilisé pour la gestion et la diffusion de messages à grande échelle.

3. Rust est un langage de programmation moderne, sûr et performant, conçu pour aider les développeurs à écrire des logiciels fiables et efficaces.

4. Un topic Kafka est une catégorie ou un canal de messages où les données sont publiées et auxquelles les consommateurs peuvent s'abonner pour lire ces messages.

destination, les horodatages de début et de fin du *NetFlow*, ainsi que le numéro du protocole.

Ces informations ont été choisies car elles représentent l'ensemble minimal de caractéristiques utilisées dans la littérature pour la détection d'attaques DDoS. Ces informations sont enregistrées au format binaire dès leur réception. Afin d'optimiser l'espace disque du serveur chargé de l'acquisition des données, les fichiers binaires sont limités à une taille approximative de 1 Go. Cette taille a été choisie expérimentalement pour trouver le bon équilibre entre le temps de capture et le temps de compression, afin d'éviter que cette rotation ne prenne trop de temps et ne sature l'espace disque du serveur de capture. Lorsque cette taille est atteinte, le fichier est compressé au format *Gzip*, qui a été choisi en raison de sa popularité et de sa disponibilité par défaut sur la distribution *GNU/Linux* utilisée sur les serveurs de capture.

Une fois que les *NetFlows* ont été archivés sur une période de 24 heures, il a été décidé d'agréger les mesures mentionnées précédemment par fenêtres d'une minute. La durée d'agrégation des fenêtres a été choisie expérimentalement pour permettre un temps de calcul raisonnable sur les données issues de la production d'OVHcloud et pour maintenir une bonne lisibilité des données issues des ensembles de données de la littérature.

Pour construire ces fenêtres d'une minute, les parties des *NetFlows* qui ont commencé avant le début de la capture ont été exclues. Il est possible qu'un *NetFlow* présent dans la capture ait commencé bien avant le début de la capture elle-même. Pour résoudre cette situation, les *NetFlows* contiennent des compteurs qui s'appliquent sur toute la durée de la capture. Par exemple, si un *NetFlow* a une durée de dix secondes avec un total de 2 000 paquets, il est impossible de déterminer si les 2 000 paquets sont répartis de manière uniforme sur les dix secondes, ou si les 1 500 premiers paquets sont arrivés à un moment $T + 2s$ et que les 500 paquets restants sont répartis uniformément entre $T + 3s$ et $T + 9s$.

Pour gérer les horodatages, une phase de prétraitement a été nécessaire afin de construire les fenêtres. Et en pratique, pour des questions de délais de transmission différents selon les points de collectes des différents *NetFlow* de la backbone, ils ne sont pas complètement ordonnés. À l'aide d'un second programme *Rust* qui prend en entrée une liste d'archives *Gzip* contenant les *NetFlows*. Ensuite, l'offset du *NetFlow* actuel a été enregistré si son horodatage de début et de fin se situait à l'intérieur de la fenêtre en cours de construction.

Enfin, à l'aide d'un troisième programme *Rust* prenant en entrée les fichiers indiquant les offsets de chaque *NetFlow* pour chaque fenêtre d'une minute, les mesures statistiques mentionnées précédemment sont calculées pour chaque fenêtre. Pour accélérer le traitement, la bibliothèque *Rayon*[5] a été utilisée pour paralléliser les trai-

5. La crate Rayon en Rust est une bibliothèque qui permet de paralléliser facilement les tâches

tements sur des itérateurs et ainsi traiter plusieurs fenêtres simultanément. Malgré l'introduction de ce parallélisme, il faut environ cinq minutes de temps de traitement, sur un machine de type *i1-18Q* de la gamme *Storage Optimized de Public cloud* d'OVHcloud, cette machine dispose de 180Go de mémoire, 32 vCores ainsi que que 4 disques NVMe de 1,9To, pour obtenir les valeurs statistiques de chaque fenêtre d'une minute. Une fois le calcul d'une fenêtre terminé, les résultats sont enregistrés sur le disque dans un fichier au format CSV [6].

3.4 Comparaison des jeux de données de la littérature et du trafic d'OVHcloud

Dans cette section, nous démontrons l'utilité des métriques que nous avons sélectionnées dans la section précédente pour comparer les jeux de données de la littérature et le trafic de production d'OVHcloud. Bien que les jeux de données soient souvent utilisés pour développer de nouvelles techniques de détection d'attaques DDoS, notre objectif ici est de mettre en évidence l'impact des attaques DDoS ainsi que du trafic nominal sur les métriques généralement utilisées par les équipes en charge de la sécurité des infrastructures réseau et des solutions de détection d'intrusions.

Dans un premier temps, nous allons montrer l'évolution de ces métriques sur du trafic issu des jeux de données de la littérature, qui est étiqueté comme légitime. Cela nous permettra de comprendre les variations naturelles du trafic nominal et d'établir une référence pour les métriques considérées. Nous utiliserons les jeux de données sélectionnés précédemment, tels que *ISCX 2012, UNSW-2015, URG'16, CIC DoS, CSE-CIC-DDoS et LITNET-2020*.

Ensuite, en utilisant ces jeux de données, nous montrerons l'évolution de ces métriques sur du trafic étiqueté comme étant constitué d'attaques DDoS volumétri-ques. Nous analyserons comment ces attaques impactent les "différentes métriques et quelles variations significatives peuvent être observées. Cette étude nous permettra de mieux comprendre les changements de comportement du trafic lorsqu'il est soumis à des attaques DDoS, ainsi que l'effet de ces attaques sur les métriques de sécurité couramment utilisées.

Enfin, nous discuterons des implications de ces résultats à l'échelle d'un fournisseurs de sevices *cloud* tel qu'OVHcloud. Les métriques que nous avons étudiées ont un impact direct sur la sécurité et la détection des attaques DDoS au sein d'une infrastructure cloud. Comprendre comment ces métriques évoluent lors d'attaques DDoS

sur plusieurs cœurs de processeur, améliorant ainsi les performances des programmes.

6. CSV (Comma-Separated Values) est un format de fichier texte utilisé pour stocker des données tabulaires sous forme de lignes, où les valeurs sont séparées par des virgules ou d'autres délimiteurs.

permet d'identifier les seuils anormaux et les signaux indicateurs d'attaques imminentes. Cela permet aux équipes de sécurité de prendre des mesures proactives pour atténuer les attaques et protéger l'infrastructure du fournisseurs de services *cloud*. Ces résultats contribueront ainsi à renforcer la résilience et la sécurité des fournisseurs de services *cloud* face aux attaques DDoS.

3.4.1 Analyse des jeux de données de la litérature

Dans cette section, nous illustrons la pertinence des métriques précédemment introduites sur des ensembles de données d'attaques DDoS issus de la littérature scientifique. L'objectif n'est pas de détailler une technique de détection spécifique, mais plutôt de mettre en valeur les données sur lesquelles les experts s'appuient pour identifier des situations de crise. Pour ce faire, nous commençons par présenter la manière dont ces métriques se manifestent sur un trafic bénin, avant de montrer comment elles peuvent être utilisées pour caractériser les attaques DDoS volumétriques. Nous discutons ensuite des implications de ces résultats à l'échelle d'un fournisseurs de services *cloud*.

La Figure 3.1 présente l'évolution des métriques que nous avons sélectionnées dans la section précédente. Dans un premier temps, nous observons cette évolution sur un dataset issu de la littérature, qui contient uniquement du trafic légitime sans attaques DDoS. Les données compilées dans la Figure 3.1 sont extraites de la journée de lundi du dataset CICDDoS2017.

En examinant la métrique caractérisant le nombre d'adresses IP sources par adresse IP destination, nous constatons une distribution aléatoire approximativement centrée autour de la valeur de 100 adresses IP sources par adresse IP destination. De plus, nous observons un effet de saisonnalité en début de journée, à partir de 00h00 GMT, correspondant à 09h00 heure locale du dataset. Cet effet de saisonnalité est également visible tôt le matin dans la métrique du nombre total de paquets par seconde, où le nombre de paquets dépasse les 10 000 par seconde au cours des premières heures pour se stabiliser à 1 000 paquets par seconde le reste de la période.

En ce qui concerne la distribution des protocoles et le calcul de l'entropie liée aux ports, l'effet de saisonnalité est encore plus marqué. Les valeurs de ces deux métriques obéissent à des distributions aléatoires centrées autour d'une valeur donnée, avec des fluctuations liées aux effets saisonniers. Par exemple, dans un contexte d'entreprise, il est possible d'observer une augmentation significative du trafic réseau pendant les heures de bureau et une diminution la nuit lorsque la plupart des utilisateurs sont hors ligne.

Ces observations soulignent l'importance de prendre en compte les variations temporelles et saisonnières dans l'analyse des métriques de trafic. Elles mettent en évidence des comportements spécifiques du trafic légitime, ce qui permet de mieux

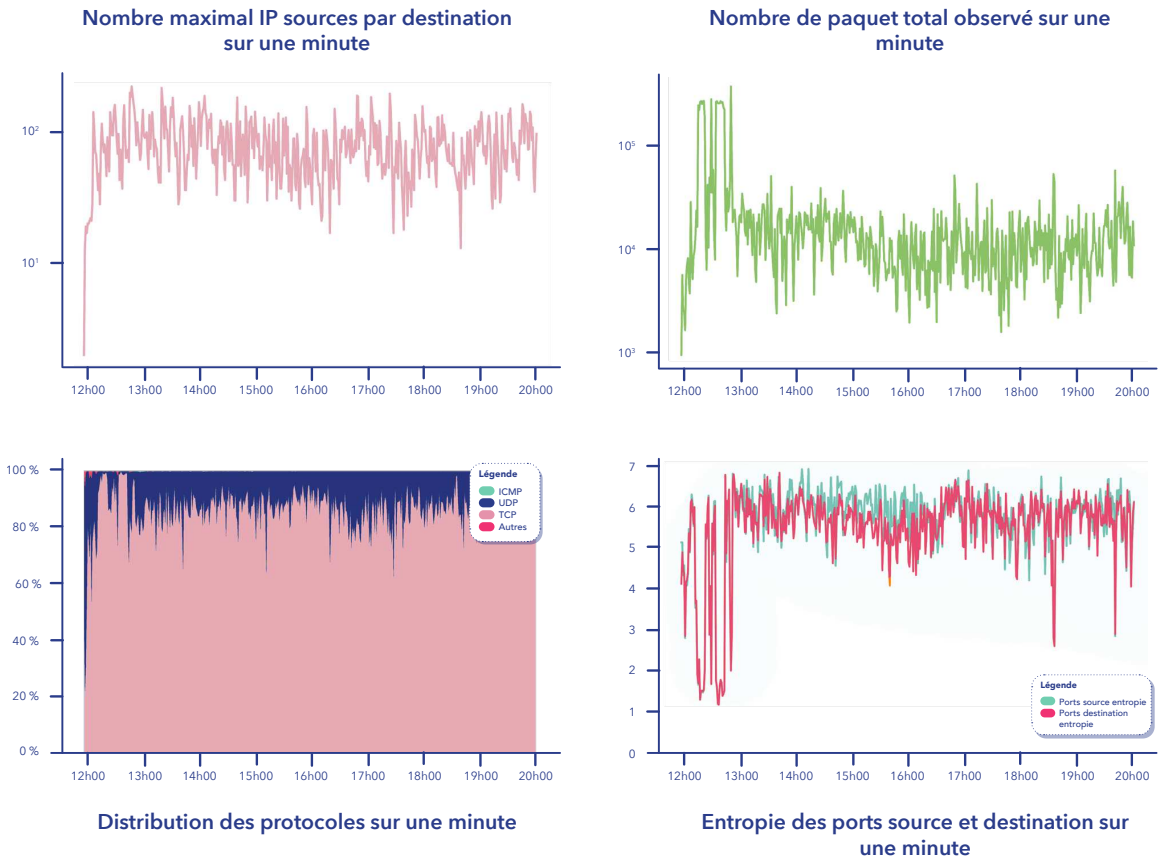

FIGURE 3.1 – CIC IDS 2017 Heures de travail Lundi.

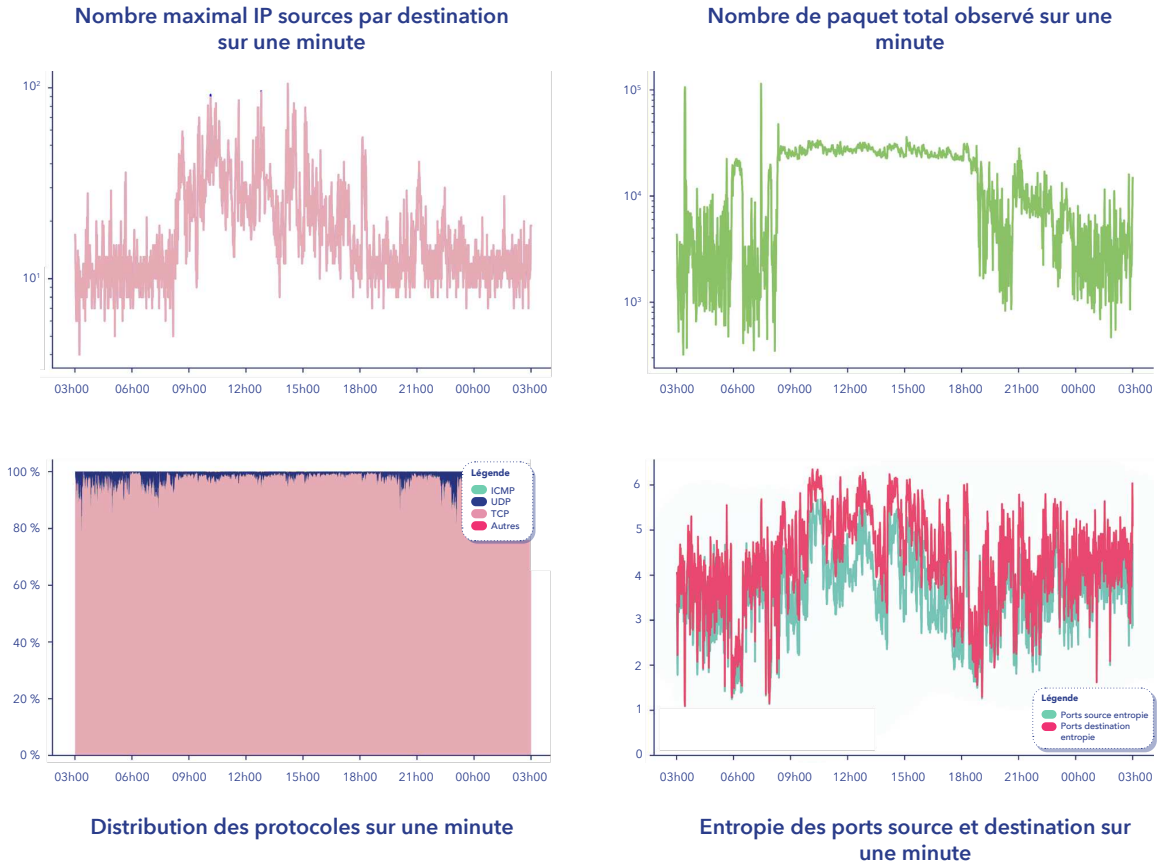

FIGURE 3.2 – ISCX IDS 2012 le 11 juin.

comprendre les caractéristiques du trafic nominal et d'établir une référence pour la détection des anomalies liées aux attaques DDoS.

Les résultats présentés dans la Figure 3.1 illustrent l'impact de l'évolution temporelle sur les métriques sélectionnées et fournissent un aperçu de l'évolution du trafic légitime dans un contexte spécifique. Ces informations sont essentielles pour le développement de techniques de détection d'attaques DDoS efficaces, qui doivent être capables de distinguer les comportements légitimes des comportements malveillants.

La journée du 11 juin, dans *ISCX IDS 2012* constitue un autre exemple de jeux de donnée ne contenant pas de trafic d'attaques DDoS. Ici aussi, les effets liés à la saisonnalité du trafic sont très marqués et facilement observables sur les métriques étudiées dans la Figure 3.2.

De 09h00 à 18h00, le trafic présente des caractéristiques différentes par rapport au trafic observé durant la nuit, pour chacune des métriques étudiées. Par exemple, le nombre maximal d'adresses IP sources par adresse IP destination est plus élevé

pendant les heures de bureau que la nuit, ce qui peut être expliqué par l'activité accrue des utilisateurs pendant cette période. De même, les ratios des protocoles montrent que le protocole TCP est davantage représenté pendant les heures de bureau, ce qui peut être attribué à une utilisation plus fréquente des applications web et des services en ligne.

En ce qui concerne le nombre total de paquets transmis par minute et sa variance, ces métriques évoluent différemment entre le jour et la nuit. Pendant les heures de bureau, le nombre de paquets transmis par minute peut être significativement plus élevé, notamment en raison de l'utilisation intensive des applications et services en ligne par les utilisateurs. Par contre, la variance du trafic peut être plus importante durant la nuit, en raison de l'absence d'une utilisation régulière et prévisible des services.

Enfin, la valeur moyenne de l'entropie des ports présente une particularité intéressante durant la période de 06h00 à 18h00. Pendant ces heures, l'entropie des ports est plus élevée, indiquant une plus grande diversité des ports utilisés, ce qui peut être attribué à une activité plus variée des utilisateurs et des applications pendant les heures de travail.

L'étude comparative des deux jeux de données, *CICDDoS2017* et *ISCX IDS 2012*, a permis de mettre en lumière des caractéristiques du trafic légitime ainsi que des variations temporelles importantes sur les métriques à l'étude. Les résultats montrent que le trafic légitime présente des effets de saisonnalité clairement identifiables avec des comportements spécifiques observés en fonction de l'usage qui est fait de l'infrastructure réseau à l'étude. Par exemple, dans le cas d'une infrastructure de type petite et moyenne entreprise (PME), on note des comportements spécifiques entre le jour et la nuit.

De même, dans le jeux de donnée *CICDDoS2017*, on observe une distribution aléatoire des adresses IP sources par adresse IP destination, une fluctuation importante du nombre total de paquets par minute et une variabilité saisonnière dans la distribution des protocoles ainsi que de l'entropie des ports sources et destination. Ces variations temporelles mettent en évidence les fluctuations naturelles du trafic dans un contexte étiqueté comme légitime.

De même, dans le jeux de donnée *ISCX IDS 2012*, les effets liés à la saisonnalité du trafic sont clairement identifiables. Pendant les heures de bureau, on observe des valeurs plus importantes du nombre d'IP sources par adresse IP destination, du ratio des protocoles, et de l'entropie des ports, comparativement à la période nocturne. Ces résultats mettent en évidence des comportements spécifiques en fonction du moment de la journée et de l'usage de l'infrastructure.

L'étude de ces deux jeux de données renforce l'importance de considérer ces variations naturelles du trafic légitime pour mieux distinguer les comportements légitimes des comportements malveillants et développer des techniques de détection d'attaques

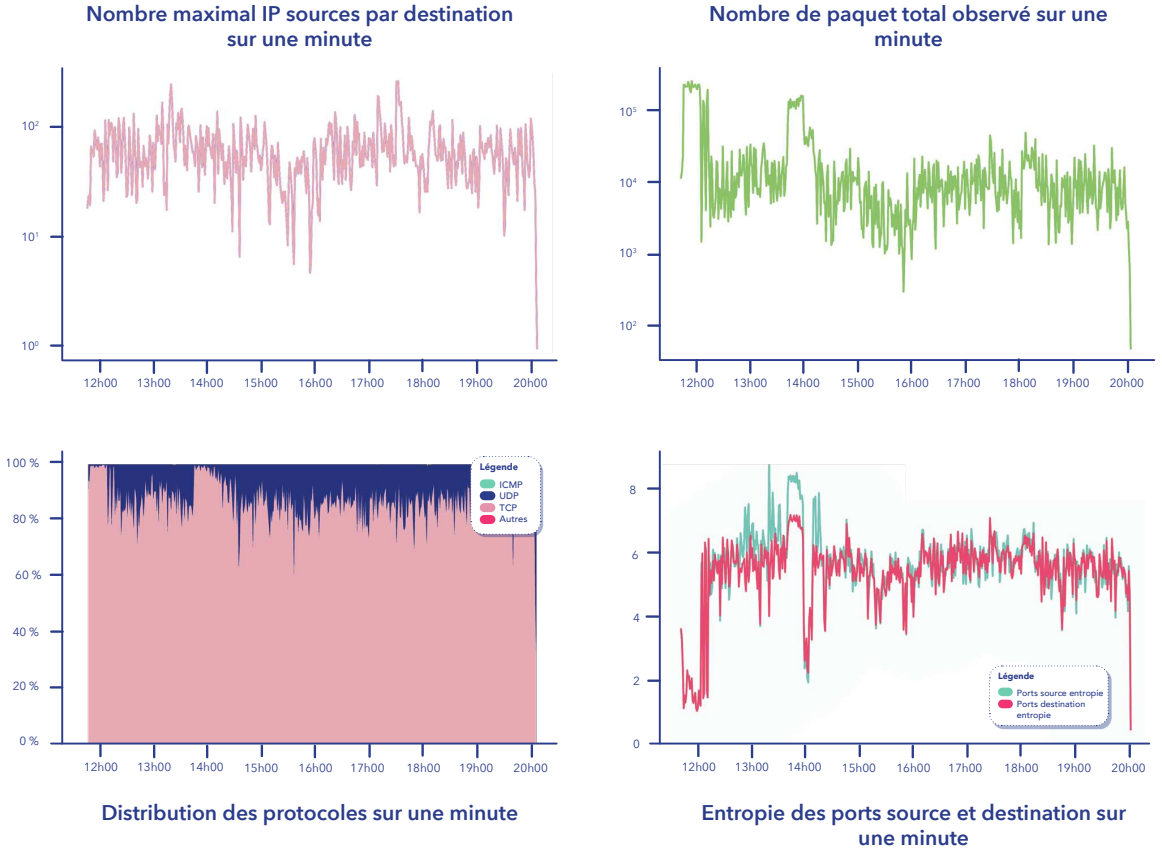

FIGURE 3.3 – CIC IDS 2017 Heures de travail du mercredi.

DDoS qui prennent en compte le contexte d'utilisation de l'infrastructure.

Pour comprendre l'évolution des métriques lorsque le trafic contient des attaques DDoS, nous avons étudié les jeux de données *CIC IDS 2017*, *CSE IDS 2018* et *CIC DDoS 2019*. La Figure 3.3 présente les données du mercredi du jeux de donnée *CIC IDS 2017*, qui contient des attaques DoS et DDoS. Les métriques que nous avons sélectionnées dans la littérature montrent de manière très identifiable ces attaques. Les attaques *Slowloris*[7] sont clairement visibles grâce à la métrique du nombre total de paquets, qui montre des pics de trafic supérieurs d'un ordre de grandeur au-dessus de la valeur moyenne. La métrique montrant l'évolution de la distribution des protocoles présente le même comportement.

Cependant, si l'on examine l'évolution de ces deux métriques lors du démarrage du pic lié à un effet saisonnier et celui lié au début d'une attaque DDoS, il peut être

7. Slowloris est une attaque DDoS qui exploite des vulnérabilités dans la gestion des connexions HTTP en maintenant un grand nombre de connexions ouvertes de manière incomplète pour saturer les ressources du serveur ciblé.

difficile de les différencier, ce qui montre que ces métriques ne sont pas suffisantes pour qualifier une attaque DDoS du trafic légitime. En revanche, la métrique de l'entropie mesurant l'entropie des ports sources diffère nettement de celle mesurant l'entropie des ports destinations lors d'une attaque DDoS. Cette métrique, combinée aux autres, constitue un marqueur important pour identifier les attaques DDoS.

Ces trois métriques, bien connues des experts du domaine, associées à la connaissance de la saisonnalité du trafic, permettent d'identifier visuellement des attaques DDoS volumétriques. Cependant, en dehors des attaques volumétriques, les attaques présentes dans ce jeu de données ne sont pas identifiables par les métriques que nous avons choisie. En effet, ces attaques se basent sur des techniques qui envoient peu de volume de trafic réseau, car elles exploitent notamment le protocole $HTTP$ pour surcharger la capacité de traitement du serveur ciblé.

En dehors des attaques DDoS qui ciblent l'infrastructure réseau d'un fournisseur, la détection d'attaques visant spécifiquement un client particulier, parfois avec des débits plus faibles en raison des ressources plus modestes de la cible, ne s'avère pas toujours réalisable à l'aide des métriques couramment utilisées. Les attaques ciblées sur des clients individuels exigent souvent des méthodes de détection plus sophistiquées et des approches personnalisées, car elles peuvent se présenter sous des formes plus subtiles et échapper aux seuils de détection traditionnels

Le jeux de donnée *CSE CIC IDS 2018* contient plusieurs attaques DDoS, dont deux ont été observées le mardi : DDoS *attacks-LOIC-HTTP* et *DDoS-LOIC-UDP*. La deuxième attaque, *DDoS-LOIC-UDP*, est clairement identifiable par les métriques que nous avons sélectionnées, comme illustré dans la Figure 3.4. Cette attaque se manifeste sur la métrique du nombre total de paquets par minute, où l'on observe un pic significatif correspondant à une variation soudaine dans la distribution des protocoles. En effet, le vecteur d'attaque utilisé pour conduire l'attaque est UDP, ce qui entraîne une fluctuation notable dans la distribution des protocoles. Cette fluctuation est également observable sur la métrique de l'entropie du port destination, ce qui permet de détecter les changements dans le trafic liés à cette attaque DDoS spécifique

La Figure 3.5 présente les mesures effectuées sur le deuxième jour du jeux de donnée *CIC DDoS 2019*. Ce jour-là, il y a eu 12 attaques DDoS, dont la majorité sont des attaques volumétriques. Les trois métriques de la littérature, à savoir le nombre total de paquets par minute, la distribution des protocoles et l'entropie des ports, identifient clairement chacune des attaques volumétriques présentes dans le jeux de donnée. Cette observation souligne la pertinence de ces métriques pour la détection d'attaques DDoS.

Cependant, la seule métrique qui ne parvient pas à qualifier les attaques DDoS selon les informations est le nombre maximal d'IP sources par destination. Ceci peut s'expliquer par le fait que ces attaques n'ont pas été conduites en utilisant un botnet

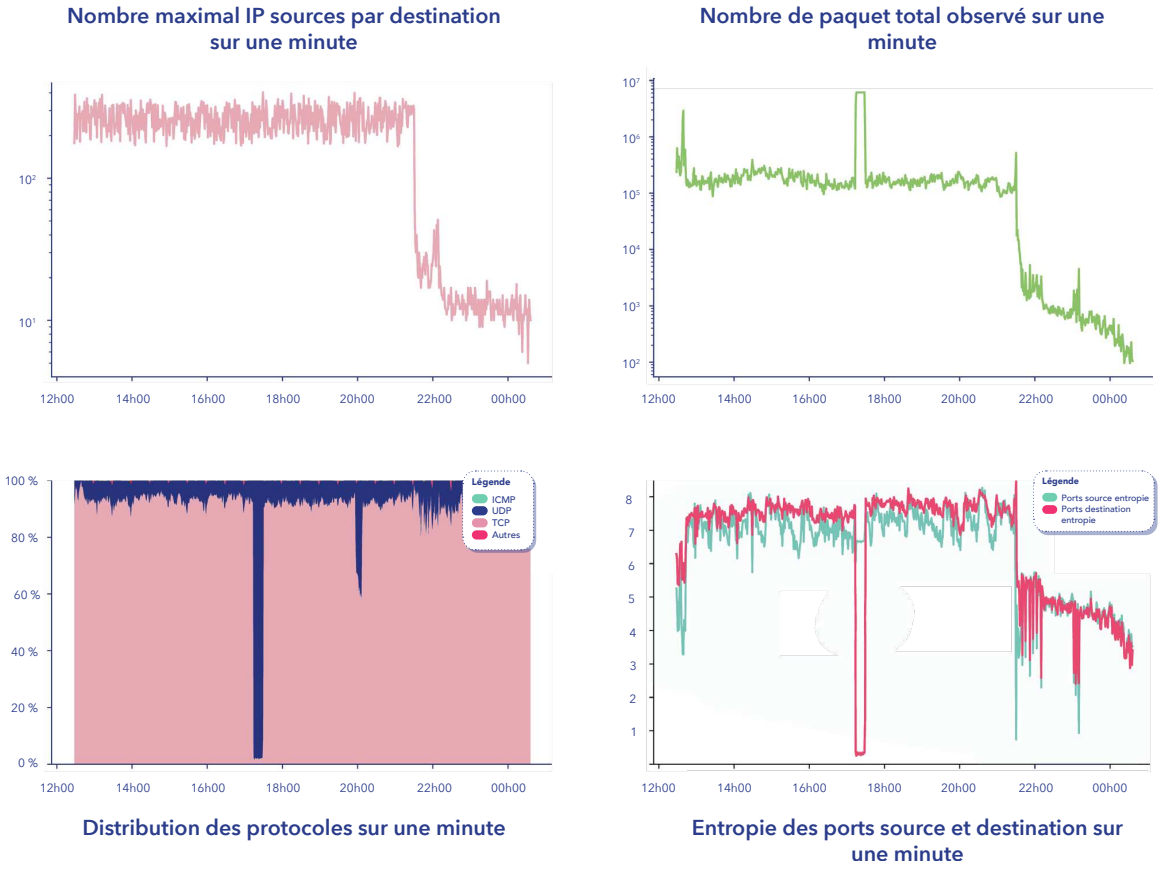

FIGURE 3.4 – CSE CIC IDS 2018 le 20-02-2018.

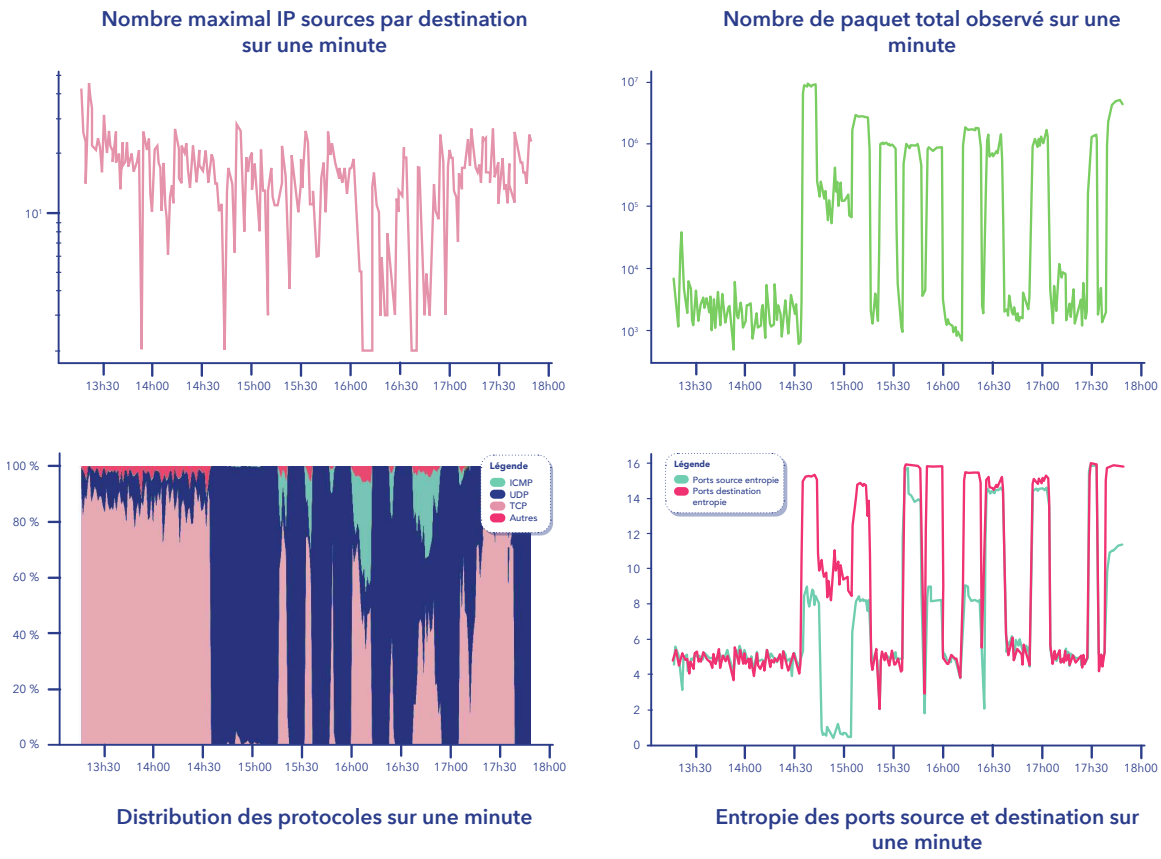

FIGURE 3.5 – CIC DDoS 2019 le 1er décembre.

et qu'une stratégie d'usurpation des IP sources n'a pas été utilisée. Dans ce contexte, les attaques DDoS sont peut-être menées par un seul attaquant ou un groupe d'attaquants avec des adresses IP sources légitimes, ce qui rend difficile la distinction entre les adresses sources malveillantes et légitimes.

Cette observation met en évidence les limitations potentielles de certaines métriques dans la détection d'attaques DDoS spécifiques. Il est important de reconnaître que les attaques DDoS peuvent être menées avec des techniques variées, et qu'une approche unique de détection peut ne pas être suffisante pour couvrir tous les scénarios d'attaques. Par conséquent, une combinaison de métriques et de techniques de détection est nécessaire pour une détection robuste des attaques DDoS dans un environnement cloud.

L'étude des jeux de données *CIC IDS 2017*, *CSE CIC IDS 2018* et *CIC DDoS 2019* a permis de mettre en évidence l'efficacité de certaines métriques pour détecter les attaques DDoS volumétriques, mais leur limitation dans l'identification des attaques DoS voire DDoS non volumétriques. Cette analyse contribue à la compréhension des différents types d'attaques et souligne l'importance de mettre en œuvre des stratégies de défense personnalisées pour lutter contre les différentes formes d'attaques dans un environnement cloud.

3.4.2 Analyse du trafic de production OVHcloud

Après avoir calculé les métriques sur les jeux de données identifiés dans la littérature, nous avons démontré leur pertinence pour détecter les attaques DoS et DDoS labellisées dans ces jeux de données. Comme ces métriques sont répandues dans la littérature, les auteurs des jeux de données *URG'16* et *LITNET-2020* ont déjà fourni le travail dans leurs travaux respectifs. Les différents graphiques produits dans ces études montrent l'efficacité de ces métriques lorsqu'elles sont combinées pour détecter les pics liés aux attaques volumétriques.

À présent que la pertinence des métriques sélectionnées à été démontrée sur les jeux de données de la littérature, leurs calculs sur le trafic réel dans un fournisseur peuvent être faits. En effet, la grande différence entre les fournisseurs et les jeux de données de la littérature réside dans le volume de trafic, ainsi que la diversité des types de trafic transitant par un fournisseurs. Par exemple, sur l'infrastructure d'OVHcloud, le nombre de paquets transitant par seconde avoisine les 2,5 milliards, tandis que dans les jeux de données que nous avons étudiés, les volumes observés n'excèdent pas les 40 000 paquets par seconde.

L'infrastructure à l'étude est composée de plusieurs centaines de milliers de serveurs hébergeant plus de 1,5 million de clients, générant un trafic de plusieurs térabits par seconde. En tant qu'infrastructure critique de production, nous avons utilisé le système d'observation du trafic déjà en place, principalement basé sur Netflow, une

technologie bien connue et documentée dans la littérature.

Cette méthode d'échantillonnage nous permet de collecter un sous-ensemble représentatif du trafic global sans nécessiter la mise en place d'un système de supervision en temps réel pour examiner chaque flux. De plus, comme le montre l'étude de Androuildakis et al. [And+06], l'efficacité de l'échantillonnage repose principalement sur le taux d'échantillonnage et non sur la méthode spécifique employée. Avec un taux d'échantillonnage approximatif de 1 pour 2000 pour le trafic entrant et de 1 pour 4000 pour le trafic sortant, et en considérant que nous observons plusieurs milliards de paquets par seconde, il est raisonnable de penser que les données obtenues ont une représentativité statistique comparable à celles extraites des jeux de données de la littérature.

Ce changement d'échelle a un impact sur la manière de calculer les métriques que nous avons présentées dans ce chapitre. Il est impossible de capturer l'intégralité du trafic à cette échelle. C'est pourquoi, pour un fournisseurs de services *cloud*, nous devons utiliser des statistiques basées sur des échantillons de trafic. L'usage d'échantillons pour effectuer des statistiques sur le trafic de production est bien documenté dans la littérature. Dans un premier temps, nous avons vérifié que l'introduction de l'échantillonnage n'affecte pas la qualité des métriques que nous avons sélectionnées. Pour cela, nous avons échantillonné un des jeux de données de la littérature que nous avons choisi d'étudier en utilisant la stratégie d'échantillonnage de *sFlow*, qui est une des méthodes de l'état de l'art.

En comparant les Figures 3.5 et 3.6, on constate qu'en dépit de l'échantillonnage à 1 pour 32, il n'y a pas de perte d'informations significative sur les valeurs des métriques calculées sur le jeux de donnée *CIC DDoS 2019*.

Ces résultats sont essentiels pour valider l'utilisation d'échantillonnage dans le calcul des métriques à grande échelle, permettant ainsi de travailler avec des volumes de trafic réalistes dans un environnement fournisseurs de services *cloud*. Cette étape est cruciale pour s'assurer que les métriques sélectionnées restent efficaces et fiables dans des conditions de trafic réel.

Pour qualifier le trafic nominal d'un fournisseurs de services *cloud*, nous avons réalisé une capture sur une durée de 12 heures, de 08h00 à 20h00, en utilisant une méthodologie constante pour tracer l'évolution des métriques sélectionnées durant cette période. Le trafic a été capturé en utilisant une méthode d'échantillonnage des paquets, qui s'avère indispensable compte tenu de l'échelle volumétrique du trafic observé.

Nous nous sommes concentrés uniquement sur les flux IPv4, car ils constituent la majorité du trafic qui transite sur la backbone d'OVHcloud à l'heure où ces lignes sont écrites. La Figure 3.7 présente donc l'évolution des métriques que nous avons observées durant notre capture de 12 heures. Le point de divergence principale que l'on peut noter entre les jeux de données moyens de la littérature et l'infrastructure

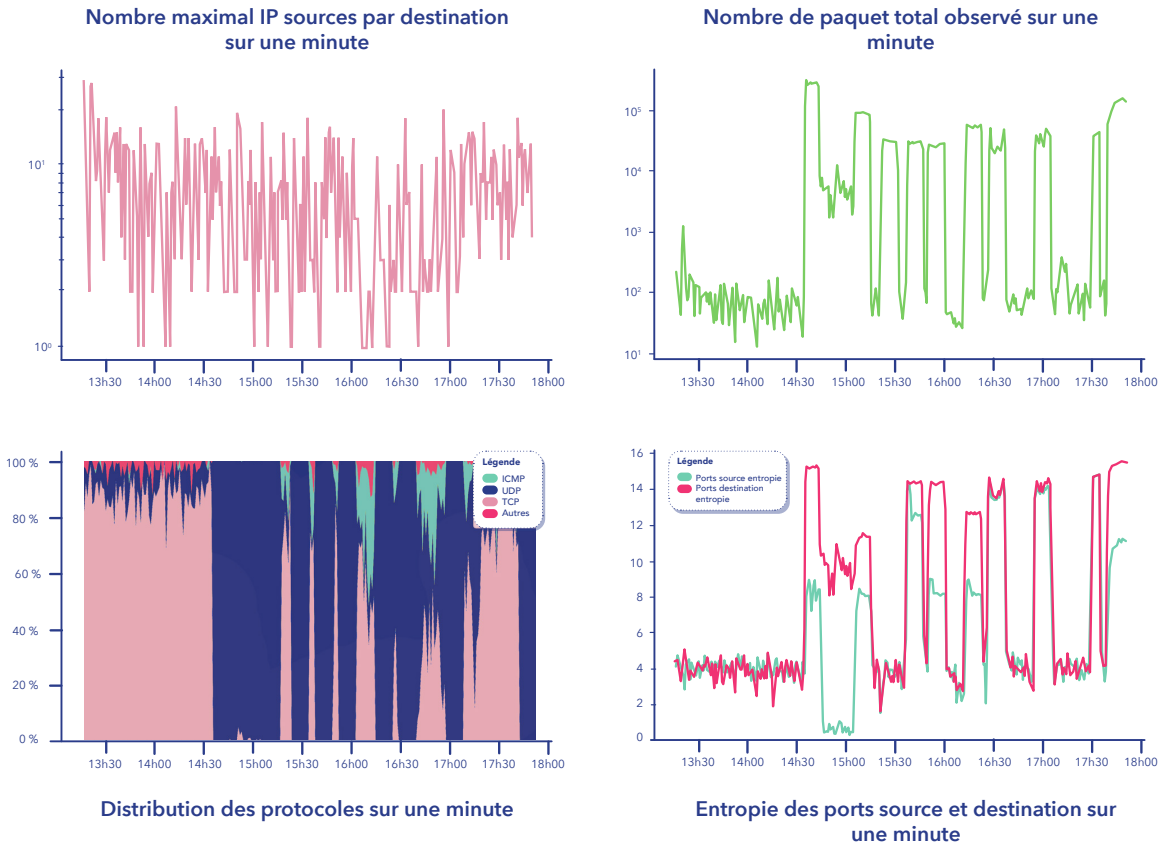

FIGURE 3.6 – CIC DDoS 2019 01-12 avec un échantillonnage à 1 pour 32

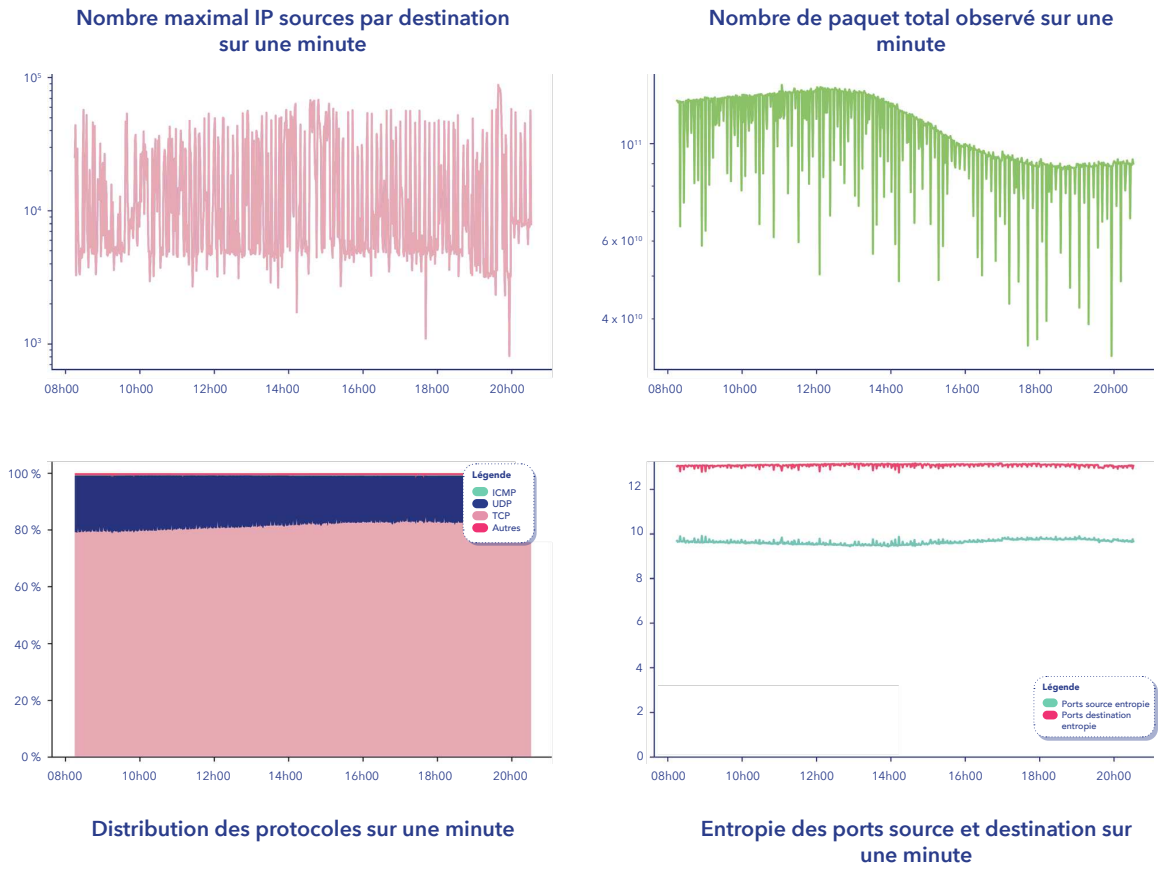

FIGURE 3.7 – Évolution des métriques observé sur la backbone de OVHcloud

d'OVHcloud est que l'évolution des métriques est homogène, malgré l'occurrence de plusieurs milliers d'attaques DDoS volumétriques, dont certaines dépassent un téra-octet par seconde, et qui ont été détectées par le système de détection déjà en place. Il n'y a donc aucun pic lié aux attaques qui soit discernable sur les courbes que nous avons obtenues, contrairement à ce que nous avons observé jusqu'à présent. Il est toutefois à noter qu'un léger effet saisonnier est observable au fur et à mesure que la journée progresse. Ces observations soulignent la nécessité d'explorer davantage les signaux faibles pour une détection précise des attaques DDoS dans un environnement cloud à grande échelle.

De plus, comme nous l'avons montré dans une étude précédente, des effets de saisonnalité sont également visibles à l'échelle d'un fournisseurs de sevices *cloud*. Cependant, ces effets, qui font partie de la vie normale d'une infrastructure fournisseurs de services *cloud*, ne semblent pas représenter d'obstacles particuliers aux techniques que l'on trouve dans la littérature dans le contexte de l'analyse réseau. Une fois qu'ils

sont pris en compte dans l'élaboration des systèmes de détection, ils peuvent être correctement gérés. Néanmoins, la quantité de trafic à laquelle nous sommes exposés présente un autre défi majeur : *la problématique des signaux faibles.*

Même lors d'une attaque DDoS volumétrique, qui peut être considérée comme importante pour l'état de l'art, au vu des quantités de trafic qui transitent normalement sur l'infrastructure d'un fournisseurs de services *cloud*, ces attaques peuvent ne représenter qu'un faible volume de la quantité de trafic observé. Cela soulève ainsi la question de l'identification des pics liés aux signaux faibles d'attaque par rapport à ceux du trafic légitime.

Dans ce contexte, il est crucial de développer des méthodes de détection capables de détecter ces signaux faibles, car ils peuvent indiquer la présence d'une attaque DDoS malgré leur relative insignifiance en termes de volume de trafic. Cette détection précise des attaques parmi un trafic massif est essentielle pour rendre l'implémentation des systèmes de détection à l'échelle d'un fournisseurs de services *cloud* efficace et pertinente. Les approches traditionnelles de détection basées uniquement sur des seuils peuvent être insuffisantes dans ce contexte, car elles risquent de manquer des attaques importantes tout en générant un grand nombre de fausses alarmes. C'est pourquoi l'utilisation de techniques avancées d'apprentissage automatique et d'analyse comportementale peut être essentielle pour détecter avec précision les attaques DDoS dans un environnement cloud à grande échelle, où les signaux faibles peuvent fournir des indices importants sur la présence d'attaques malveillantes.

Pour mieux comprendre le phénomène des signaux faibles dans un contexte de trafic massif, nous avons réalisé l'expérience suivante. Nous avons multiplié par 100 le trafic légitime représenté dans la Figure 3.3 et avons ajouté une instance du trafic d'attaques DDoS illustrée dans la Figure 3.5 pour générer la Figure.

Dans le but de mieux appréhender la complexité des signaux faibles dans un environnement caractérisé par un trafic massif, j'ai mené l'expérience suivante. Tout d'abord, j'ai utilisé les données de la journée du mercredi provenant de l'ensemble de données *CIC IDS 2017* pour représenter le trafic réseau légitime. La Figure 3.3 présente une visualisation de ce trafic, qui peut être assimilé aux flux normaux au sein d'un réseau, englobant notamment les requêtes de navigation sur Internet, les transferts de fichiers, les demandes de courrier électronique, et autres interactions similaires.

Ensuite, dans le cadre de cette expérience, j'ai procédé à une multiplication de la quantité de trafic légitime par un coefficient de 100. En d'autres termes, j'ai artificiellement amplifié le volume de données relatives au trafic légitime. Cette manipulation a eu pour objectif de simuler une situation de trafic réseau excessivement dense, telle que celle qui peut être rencontrée dans le contexte des fournisseurs de services cloud, par exemple.

Outre cette augmentation substantielle du trafic légitime, j'ai également intégré

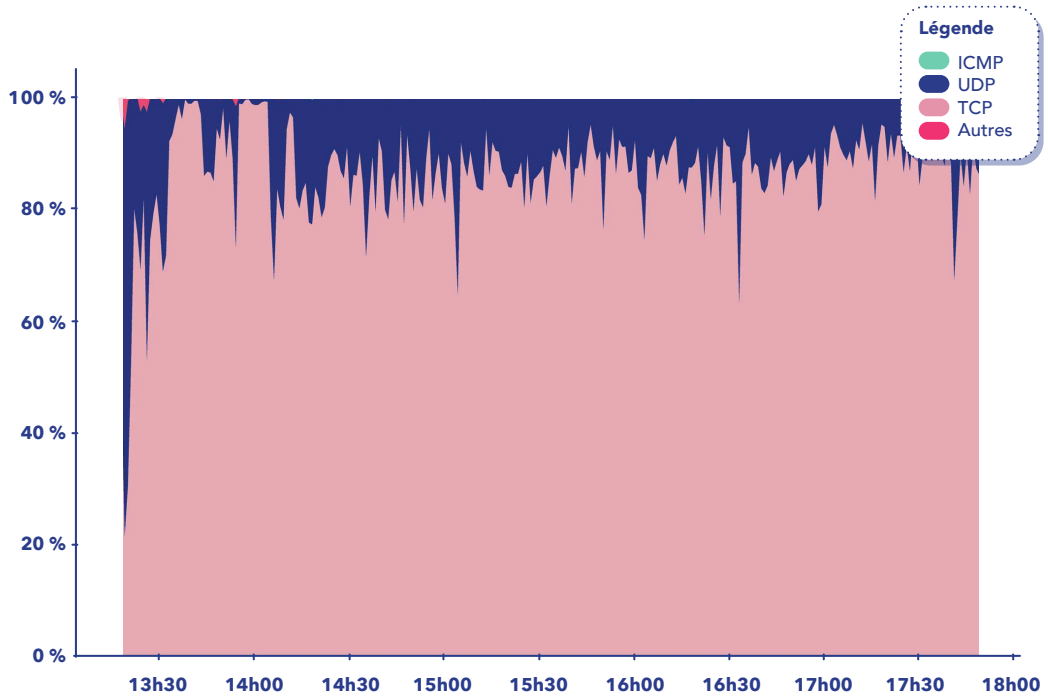

FIGURE 3.8 – Distribution des protocoles observé lors de la fusion des attaques CIC DDoS 2019 et du trafic bénin CIC DDoS 2017 avec un ratio de 1/100.

dans l'expérience du trafic issu du jeu de données *CIC DDoS 2019*, qui comprend des données d'attaques DDoS. En combinant ce trafic d'attaques DDoS avec le trafic légitime massif, mon objectif était de démontrer que même les attaques DDoS de grande envergure peuvent se présenter comme des signaux faibles dans cet environnement.

Les signaux faibles en question peuvent englober divers éléments tels que des modèles de comportement suspects, des anomalies de trafic, des pics d'activité soudains, ou d'autres indicateurs subtils. La pertinence de ces signaux faibles réside dans leur capacité à être perçus comme des indices révélateurs d'une attaque DDoS, au milieu du tumulte généré par le trafic légitime massif. En d'autres termes, cette expérience a démontré que même lorsque des attaques DDoS de grande ampleur se produisent, elles peuvent être discrètes au point de ressembler à des signaux faibles, mettant ainsi en évidence la nécessité pour les systèmes de détection d'attaques de développer des techniques capables de les identifier au sein de ce contexte particulièrement bruyant.

Dans ce scénario, nous pouvons clairement observer que les informations liées au trafic d'attaques DDoS se retrouvent diluées dans le volume total de trafic. Les attaques volumétriques deviennent ainsi des signaux faibles, submergées par le volume massif de trafic légitime.

De même, comme montré dans les études conduites par Damasevicius et al. [Dam+20]

et Macia-Fernandez et al. [Mac+18], il est possible d'observer que l'impact des attaques DDoS volumétriques reste visible sur les métriques que nous avons sélectionnées, mais il s'amenuise au fur et à mesure que l'échelle de l'infrastructure augmente. Ces analyses exacerbent les défis que représente la détection de ces attaques dans des infrastructures de type fournisseurs de services *cloud*. La première problématique réside dans les signaux faibles que ces attaques génèrent par rapport à la quantité énorme de trafic qui circule sur ce type d'infrastructures. Il est clair que les métriques traditionnelles fonctionnent très bien sur des infrastructures de taille modérée, mais peuvent s'avérer insuffisantes dans le cadre d'infrastructures de grande envergure.

3.5 Conclusion

En conclusion, notre étude a mis en évidence l'importance de la détection des attaques DDoS dans les infrastructures cloud, en particulier dans le contexte des fournisseurs de sevices *cloud*. Nous avons examiné différentes métriques traditionnellement utilisées dans la littérature pour détecter ces attaques et avons démontré leur pertinence sur des jeux de données provenant de la littérature, tels que *CIC IDS 2017* et *ISCX IDS 2012*. Nous avons également analysé le trafic réel d'un fournisseurs de services *cloud*, en l'occurrence OVHcloud, pour mettre en évidence les défis supplémentaires liés à l'échelle et aux signaux faibles que les attaques DDoS peuvent représenter dans un tel environnement.

Nos résultats montrent que les métriques traditionnelles sont efficaces pour détecter les attaques DDoS volumétriques dans des infrastructures de taille modérée, mais qu'elles peuvent être insuffisantes pour détecter les attaques dans des environnements de grande échelle. Les attaques volumétriques peuvent devenir des signaux faibles, submergées par le volume de trafic légitime circulant sur les fournisseurs de services *cloud*. Cela met en évidence la nécessité de développer des méthodes de détection plus avancées qui tiennent compte de l'échelle et des spécificités des infrastructures cloud.

Étant donné que les métriques traditionnelles se sont avérées efficaces pour la caractérisation du trafic au sein de fournisseurs de services *cloud* tels qu'OVHcloud, je vous invite à approfondir ce sujet dans le chapitre 4. Dans ce contexte, nous allons explorer comment ces métriques ont été utilisées en tant que paramètres d'entrée dans le cadre d'une preuve de concept visant à créer un générateur de trafic statistiquement réaliste. Cette approche vise à reproduire de manière précise le trafic observé au sein de l'infrastructure d'OVHcloud. Cette démarche nous permettra de mieux comprendre et de modéliser le comportement du trafic, contribuant ainsi à l'amélioration des systèmes et des solutions destinés à ce type d'environnement.

Chapitre 4

Générateur de trafic représentatif des fournisseurs de services cloud

Sommaire

4.1 Motivations

Dans le chapitre précédent, nous avons examiné la possibilité de caractériser le trafic d'un fournisseur en utilisant des métriques établies dans l'état de l'art. En utilisant cette caractérisation statistique comme point de départ, je propose de concevoir un générateur de trafic qui utilise ce « modèle » statistique comme entrée pour générer du trafic correspondant. L'idée d'un générateur de trafic me semble être une solution acceptable pour résoudre le problème de l'indisponibilité de jeux de données contenant du trafic réel, comme nous l'avons mentionné précédemment. De plus, l'utilisation d'un générateur nous permettrait de réaliser d'importantes économies en termes de ressources. En effet, la capture de trafic, même sur une période de quelques minutes, devient un défi considérable lorsque le trafic atteint des débits de plusieurs téraoctets par seconde. En plus des problèmes liés au stockage, cela exerce également une pression supplémentaire sur les équipements réseau, qui doivent dupliquer le trafic qu'ils observent pour le transmettre à l'appareil de capture.

Par conséquent, l'utilisation d'un générateur de trafic nous permettrait de nous concentrer sur la création de scénarios d'attaques personnalisés en fonction de nos besoins spécifiques du moment. En collaboration avec des ingénieurs d'OVHcloud et mes encadrants de l'équipe 2XS, j'ai entrepris de réaliser une preuve de concept d'un générateur de trafic basé sur les « modèles » statistiques précédemment établis (3). Cette initiative vise à résoudre plusieurs problèmes majeurs auxquels nous sommes confrontés, notamment l'indisponibilité de données réelles, les contraintes liées au volume élevé de trafic, les problèmes de stockage et les pressions sur les équipements réseau. De plus, un générateur de trafic est préférable à la fourniture de captures de trafic réseau lorsqu'il s'agit de respecter la réglementation en matière de confidentialité des données. Les captures de trafic peuvent contenir des informations sensibles et exposer des données privées, ce qui peut entraîner des problèmes de conformité avec les réglementations sur la protection de la vie privée. Un générateur de trafic peut être configuré pour générer du trafic fictif, anonyme et exempt de données personnelles, simplifiant ainsi la conformité tout en préservant la vie privée des utilisateurs. Cette approche élimine la nécessité d'obtenir le consentement des utilisateurs pour la collecte de données sensibles qui peut être complexe dans le cadre d'une entreprise comme OVHcloud et donc réduit la complexité administrative liée à la gestion des données personnelles .En fin de compte, un générateur de trafic semble être une solution acceptable pour nos besoins, nous permettant de personnaliser nos scénarios d'attaques tout en économisant des ressources précieuses.

4.2 Les générateurs de trafic

La littérature scientifique foisonne de générateurs de trafic réseau aux fonctionnalités diverses. Parmi eux, ID2T [Cor+15] se distingue par sa capacité à créer des simulations de trafic inter-domaine d'une grande authenticité. Cet outil s'avère particulièrement précieux pour l'évaluation des performances des systèmes de sécurité réseau et des systèmes de détection d'intrusion. Son originalité réside dans son utilisation de modèles de trafic basés sur des données réelles, permettant ainsi de générer un trafic Internet véritable. Les utilisateurs disposent d'une marge de personnalisation considérable, avec la possibilité de configurer diverses caractéristiques du trafic, allant du choix du protocole à la distribution du débit, parmi d'autres paramètres. ID2T s'érige en un instrument de choix pour la création de scénarios de trafic réseau d'une grande pertinence et d'un réalisme convaincant. Mais le fait qu'il faille donner un fichier `PCAP` contenant le trafic de « bruit de fond » en entrée, ainsi que le format de sortie soit également un `PCAP`, pose un problème de ressource à notre échelle. En effet, la problématique réside dans le fait que la génération de sorties sous forme de fichiers PCAP complets – contenant tous les paquets réseau – engendre d'énormes

volumes de données, ce qui pose des défis en matière de stockage pour les fournisseurs traitant d'importants débits de trafic. Cette approche peut entraîner des coûts élevés de stockage, des difficultés de gestion.

MoonGen [Emm+15], pour sa part, se positionne comme un générateur de trafic open source spécialement conçu pour tester et évaluer les performances de cartes réseau. Son point fort réside dans son optimisation poussée et son support varié de modèles de trafic, notamment la génération de paquets à des débits exceptionnellement élevés. Cet outil se révèle inestimable pour l'évaluation des performances de cartes réseau opérant à des débits de 10G, 25G, 40G, voire 100G, et s'avère également essentiel dans le développement de pilotes de cartes réseau. D'après notre expérience au sein de la société OVHcloud, MoonGen s'avère d'une grande utilité. Malgré seulement quatre machines, nous avons la capacité de générer plusieurs dizaines de millions de paquets par seconde par serveurs dédiés équipés de processeurs Intel Xeon CPU E3-1230 v5 cadencés à 3,40 GHz, avec 64 Go de mémoire vive et des cartes réseau de la famille Mellanox ConnectX-3 ou ConnectX-4 à double port 40 Gb/s. En ayant recours à un script Lua pour la personnalisation des paquets à émettre. Cela inclut la possibilité de générer des paquets invalides, une caractéristique peu fréquente, étant donné que la majorité des utilisateurs ne ciblent pas ce type de trafic.

TRex [AM+16], un générateur de trafic open source issu des laboratoires de développement de Cisco, se démarque par sa polyvalence. Il peut simuler un trafic conforme aux normes $Ethernet$, $IPv4$, $IPv6$, TCP, UDP, et bien plus encore à des débits élevés. L'outil est en mesure de simuler un grand nombre d'utilisateurs et de créer des millions de sessions, en faisant un outil inestimable pour les tests de performance à grande échelle des équipements réseau. En somme, il s'agit d'un générateur de trafic particulièrement prisé pour l'évaluation des performances des réseaux. Trex se présente comme une option robuste en vue de finaliser notre preuve de concept et d'en faire un outil accessible à l'ensemble des utilisateurs.

Pktgen [Ols05], un générateur de trafic open source, repose sur le kit de développement de plan de données (DPDK) et se spécialise dans les tests de performances des cartes réseau et des pilotes DPDK. L'outil offre une granularité de contrôle sur la génération de trafic, permettant aux utilisateurs de configurer en détail les caractéristiques du trafic au niveau des paquets. Pktgen est un composant essentiel des tests de performances des cartes réseau et des pilotes réseaux, apportant une souplesse de configuration très appréciée.

Warp17 [Rod], enfin, est un générateur de trafic open source spécialement conçu pour la validation des réseaux et des applications réseau. Basé sur le *framework* de test $FD.io$ (*Fast Data I/O*), il prend en charge la génération de trafic TCP/UDP, ainsi que la capture de trafic. Warp17 est l'outil utilisé pour l'évaluation des performances des applications réseau, la vérification de la stabilité du réseau et l'évaluation des dispositifs de sécurité réseau. Sa polyvalence lui confère une pertinence dans une

multitude de contextes d'évaluation et de test en réseau.

Les générateurs de trafic mentionnés précédemment, bien qu'extrêmement utiles dans divers contextes, nécessitent des adaptations significatives pour répondre à nos besoins spécifiques. Par exemple, certains de ces générateurs requièrent des captures de trafic nominal en entrée pour générer un trafic de « bruit de fond ». Cependant, cette approche pose un défi de taille, car effectuer une capture du bruit de fond d'OVHcloud s'avère complexe. De plus, d'autres générateurs exigent une description précise des comportements des utilisateurs, ce qui est également problématique, étant donné que nous n'avons qu'une vue partielle du trafic transitant dans notre infrastructure, en raison de notre travail sur des échantillons. Il convient de noter que certains générateurs non mentionnés ici sont de nature propriétaire ou nécessitent des équipements matériels spécifiques. Étant donné les limitations inhérentes à leur adaptation à nos besoins, j'ai choisi de ne pas les aborder dans cette étude.

Ainsi, l'adaptation de ces générateurs pour qu'ils correspondent à nos exigences spécifiques représente un travail substantiel. Cela nécessite une analyse minutieuse et une transformation significative de ces outils pour les aligner sur nos objectifs. En réponse à ce défi, nous avons opté pour une approche en deux temps. Tout d'abord, nous avons entrepris de réaliser une preuve de concept visant à valider nos hypothèses et à tester la faisabilité de notre démarche. Cette étape préliminaire nous a permis de mieux comprendre les ajustements nécessaires pour adapter ces générateurs à notre environnement.

4.3 Implémentation d'une preuve de concept d'un générateur

Il s'agit donc d'un générateur de trafic réseau conçu pour simuler à la fois du trafic nominal et du trafic d'attaque. Le but principal de ce générateur de trafic est de créer des données de trafic synthétiques qui peuvent être utilisées pour évaluer l'efficacité des mécanismes de détection des attaques DDoS et pour mener des expériences de recherche.

Le code est principalement écrit en Rust, un langage de programmation moderne qui offre des performances élevées et une sécurité de mémoire robuste. Il utilise plusieurs bibliothèques tierces pour gérer des tâches telles que la création de fichiers $PCAP$, la génération de nombres aléatoires, la manipulation de données au format $TOML$, et la gestion du temps. Bien que le format de sortie soit au format $PCAP$, l'échantillonnage du trafic généré permet de générer des fichiers de sortie considérablement moins volumineux que ceux produits par ID2T, par exemple.

Le générateur de trafic repose sur un certain nombre de structures de données personnalisées, telles que `ValueProbability`, `PortProbability`, `PayloadProbability`,

Algorithm 1 Génération de trafic pour la détection d'attaques DDoS

```
Lire la configuration depuis le fichier config.toml
Initialiser une structure de données pour stocker les statistiques horaires
Créer un canal (channel) pour la communication entre les threads
for chaque statistique horaire dans la configuration do
    Générer un ensemble d'adresses IP source uniques
    Générer un ensemble d'adresses IP destination uniques
    Calculer le décalage temporel en fonction de l'heure
    Générer des paquets en utilisant les distributions probabilistes
    Envoyer les paquets générés au canal
end for
Créer des fichiers PCAP distincts pour chaque heure
while il y a des paquets à écrire dans le canal do
    Récupérer les paquets du canal
    Écrire les paquets dans le fichier PCAP correspondant à l'heure
end while
Fermer proprement le canal
```

`ProtocolProbability`, et `HourlyStat`, pour stocker des informations importantes sur la configuration du trafic généré. Ces structures de données sont utilisées pour spécifier des paramètres tels que les adresses IP source/destination, les ports source/-destination, les protocoles réseau, et les tailles de la charge utile.

Une partie centrale du code est la fonction `generate_packets()`, qui génère des paquets réseau en fonction des statistiques horaires spécifiées dans la configuration. Cette fonction utilise des distributions probabilistes pour décider des caractéristiques de chaque paquet, telles que les adresses IP, les ports, les protocoles, et les tailles de la charge utile. Les paquets générés sont ensuite envoyés à un canal pour être écrits dans des fichiers $PCAP$.

La lecture de la configuration se fait à partir d'un fichier $TOML$ externe. Ce fichier de configuration contient des informations cruciales pour la génération du trafic, telles que la décision de générer ou non des charges utiles, le nom du fichier de sortie $PCAP$, et les statistiques horaires détaillées.

Un thread distinct est utilisé pour gérer la génération de paquets, ce qui permet de générer des données de trafic pour chaque heure spécifiée dans la configuration. Une fois les paquets générés, ils sont écrits dans des fichiers $PCAP$ distincts, un pour chaque heure, afin de bien séparer les données de trafic simulé. Un mécanisme de gestion de fichier est mis en place pour assurer que les fichiers $PCAP$ sont créés et gérés correctement.

Enfin, une fois la génération de paquets terminée, le canal est fermé, signalant

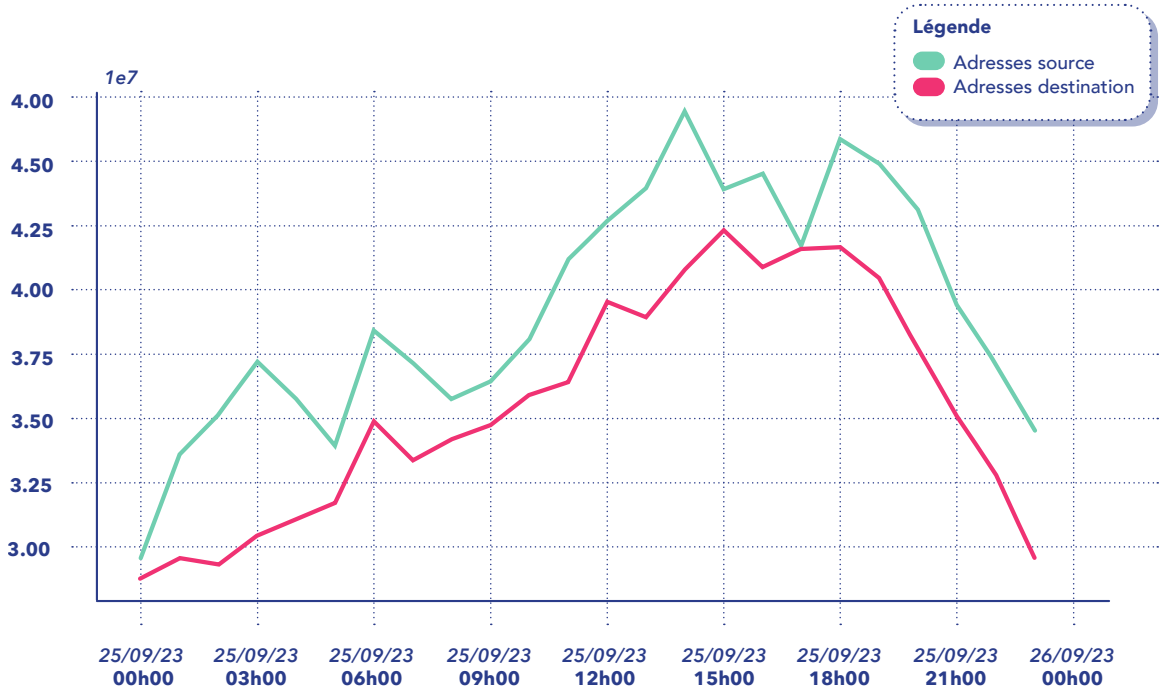

FIGURE 4.1 – Évolution des adresses IP source et destination

ainsi au *thread* d'écriture qu'il n'y a plus de paquets à générer. Cela garantit une sortie propre du générateur de trafic.

4.4 Utilisation du générateur de trafic

Une fois la preuve de concept du générateur réalisée, j'ai entrepris de générer le trafic observé sur l'infrastructure globale d'OVHcloud pour la période du 25/09/2023 au 26/09/2023. Pour ce faire, la première étape a été de réaliser une description statistique du trafic observé. À cette fin, j'ai utilisé une infrastructure d'observation du trafic récemment mise en place chez OVHcloud, reposant sur la solution *ApacheDruid*. Cette infrastructure permet de naviguer dans une base de données qui stocke l'ensemble des netflows émis par les divers équipements réseau sur une fenêtre glissante de trois semaines.

J'ai donc construit la description statistique en extrayant les données stockées à l'aide de requêtes *SQL* et en générant un fichier au format *CSV*. Par la suite, j'ai utilisé ce fichier dans un script écrit en *Python*3 en utilisant la bibliothèque *Pandas* pour générer les graphiques suivants.

La Figure 4.1 présente l'évolution du nombre distinct d'adresses IP sources et

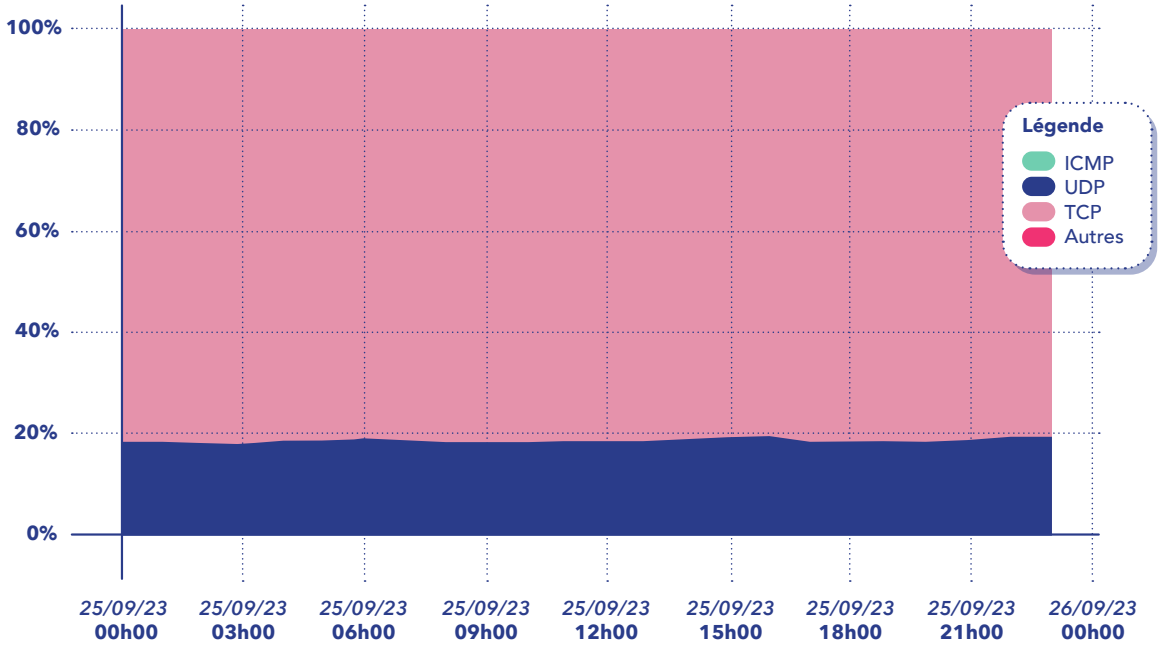

FIGURE 4.2 – Représentation en pourcentages des protocoles ICMP, UDP, TCP et Autres

d'adresses IP de destination observées au cours de cette période de 24 heures. Comme nous l'avons mentionné dans le chapitre précédent, nous avons observé un effet de saisonnalité, expliqué par les services hébergés par OVHcloud et ses clients.

La Figure 4.2 représente la répartition des protocoles $ICMP$, UDP, TCP et *Autres*. J'ai maintenu la même méthode de représentation que celle utilisée dans le chapitre précédent, et nous pouvons tirer des conclusions similaires.

La Figure 4.3 montre l'évolution du nombre total de paquets. Ici encore, nous constatons les mêmes conclusions, notamment la corrélation entre l'évolution du nombre d'adresses observées et le nombre de paquets observés.

Grâce aux statistiques extraites lors de la section précédente, ainsi qu'à la méthodologie que j'ai présentée au chapitre 3, j'ai pu élaborer le fichier de configuration, comme illustré par 4.1.

Ce fichier permet de générer un trafic représentatif de la production d'OVHcloud. La Figure 4.4 reprend l'évolution du nombre d'adresses IP source et destination distinctes, mais cette fois dans le trafic généré. Nous pouvons observer que cette évolution suit la même tendance que celle issue des chiffres de la production.

Quant à la Figure 4.5, elle présente la répartition des protocoles $ICMP$, UDP, TCP et *Autres*. Bien que cette figure illustre un comportement similaire à celui que nous observons dans le trafic de production, il est à noter que la répartition des

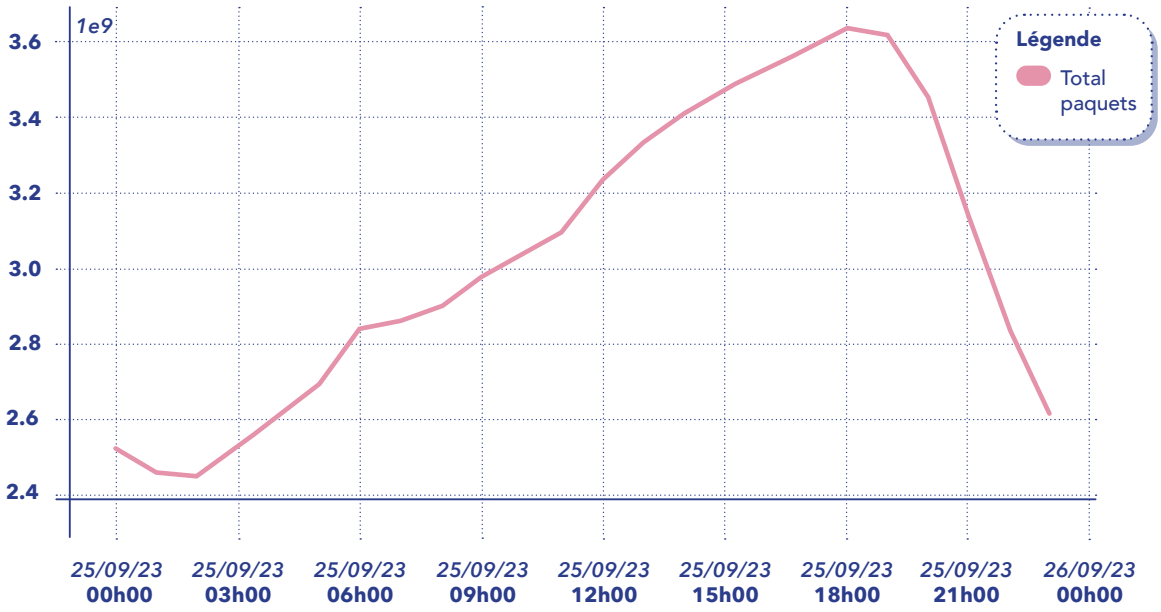

FIGURE 4.3 – Représentation de l'évolution du nombre de paquets observés

protocoles *Autres* diffère légèrement. Cette différence s'explique par la méthode de génération qui repose sur un calcul de probabilité pour déterminer à quel protocole appartient chaque paquet. C'est pourquoi les protocoles *Autres* sont sur-représentés.

En ce qui concerne la Figure 4.6, qui montre l'évolution du nombre de paquets au fil du temps, elle reflète la même évolution que celle observée dans le trafic de production.

Cette démonstration d'un générateur de trafic, conçu pour refléter les caractéristiques statistiques du trafic de production d'OVHcloud, constitue une étape essentielle dans la création rapide de jeux de données. À notre connaissance, la plupart des générateurs de trafic ne permettent pas d'effectuer une modélisation statistique du trafic qu'ils doivent reproduire. Par exemple, ID2T, qui se rapproche le plus de notre approche, nécessite un fichier *PCAP* du trafic nominal à générer, mais cette méthode est inapplicable dans notre cas en raison des débits élevés que nous observons, dépassant 7 To/s en moyenne sur notre « backbone ». La capture d'un tel fichier de paramétrage serait impossible à réaliser sans perturber notre infrastructure. En effet, même sur nos centres de données les moins sollicités, les débits dépassent les vitesses d'écriture des SSD les plus performants que nous avons à notre disposition. C'est en grande partie cette problématique qui a motivé les choix de conception de cette preuve de concept de générateur de trafic, nous permettant de contourner ces contraintes opérationnelles critiques tout en répondant à nos besoins spécifiques.

```
generate_payload = false
pcap_file = "output.pcap"
gen_date = "2023-09-25T00:00:00"

[hourly_stats]
[[stats]]
hour = 0
num_src_ips = 29467877
num_dst_ips = 28791397
total_packets = 2521981195
payload_size_distribution = [
 { value = [100, 1000], probability = 1 }
]
src_ports_distribution = [
 { value = [0, 65535], probability = 1 }
]
dst_ports_distribution = [
 { value = [0, 65535], probability = 1 }
]
protocols_distribution = [
  { value = 17, probability = 0.05 },
  { value = 6, probability = 0.7 },
  { value = 1, probability = 0.2 },
  { value = 0, probability = 0.05 }
]
```

Listing 4.1 – Début d'un fichier de configuration

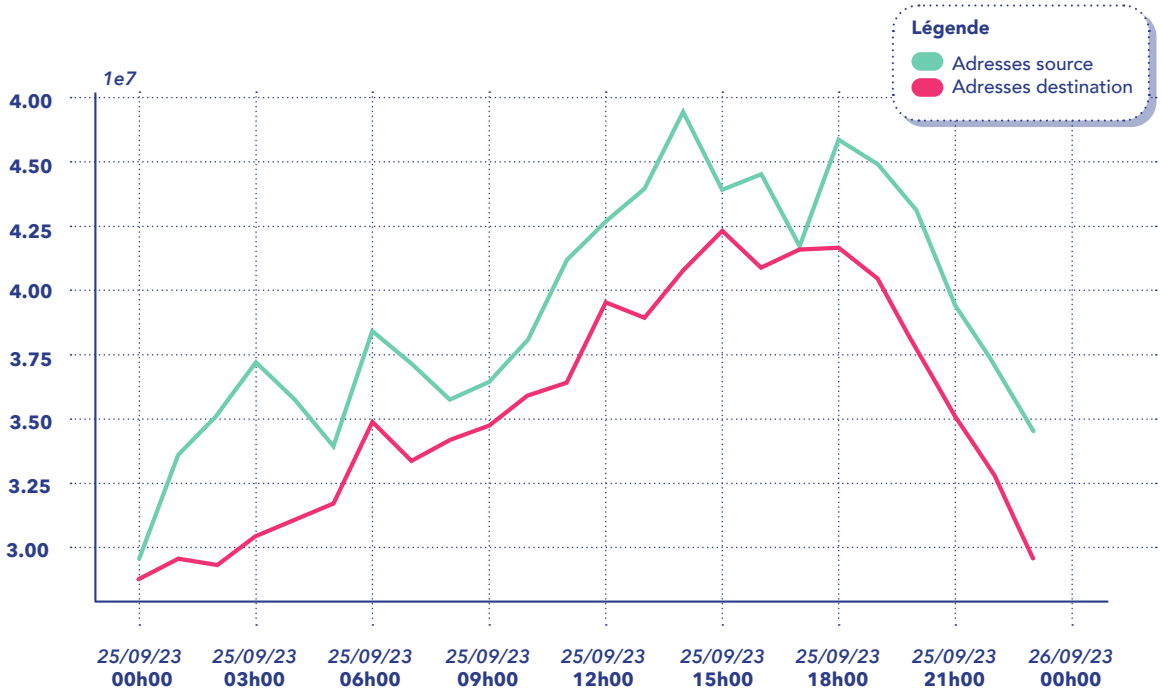

FIGURE 4.4 – Évolution des adresses IP source et destination (générateur)

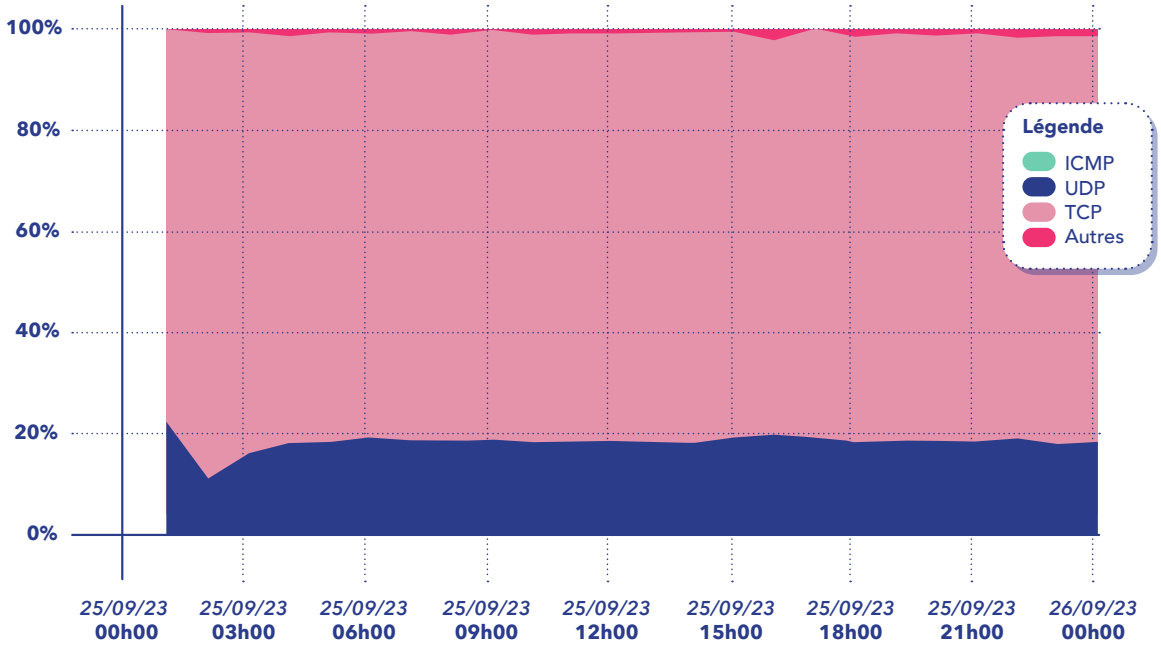

FIGURE 4.5 – Représentation en pourcentage des protocoles ICMP, UDP, TCP et Autres (générateur)

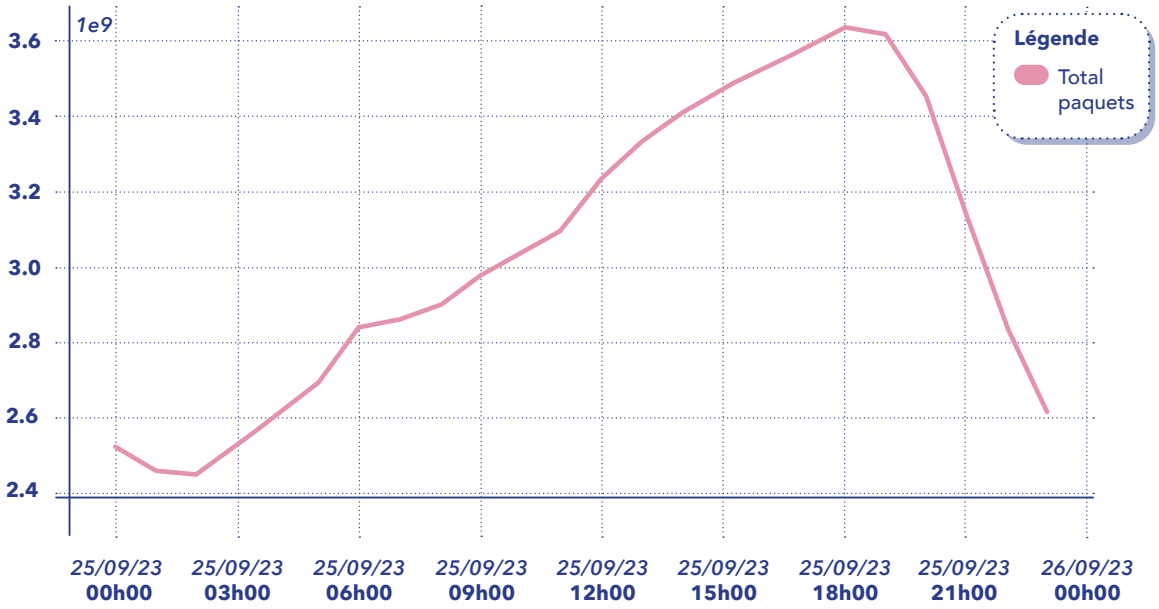

FIGURE 4.6 – Représentation de l'évolution du nombre de paquets observés (générateur)

4.5 Conclusion

Au fil de ce chapitre, nous avons pu constater que, bien qu'il ne soit encore qu'à l'état de preuve de concept, ce générateur représente un outil doté d'un potentiel considérable. En effet, en une seule journée de travail, j'ai réussi à construire un jeu de données déjà relativement fidèle à la réalité observée sur l'infrastructure d'OVHcloud. Bien que le temps de génération puisse sembler long, il est important de noter que c'était la première fois que je me lançais dans ce type de procédure, en utilisant des outils tout juste déployés en production chez OVHcloud. De plus, ce délai est nettement plus court que celui requis pour la procédure de traitement des données qui a servi à produire le matériel nécessaire aux travaux du chapitre 4.

Par ailleurs, la méthode que j'ai élaborée pour mener à bien les travaux de ce chapitre exige d'importantes ressources à savoir, un serveur dédié de la gamme HGR-SDS-1 d'OVHcloud, équipé d'un processeur Intel Xeon Gold 6242R, avec 96 Go de mémoire DDR4 ECC à 2933 MHz, et un système de stockage de 6 disques SSD SAS de 3,84 To en mode RAID logiciel. J'ai réussi à diviser par deux l'espace de stockage requis entre ces deux procédures, ce qui est loin d'être négligeable étant donné que les données générées pour ces travaux demandent tout de même près de un téraoctet d'espace disque.

De plus, la capture de données à partir des équipements de production nécessitait

auparavant près de quatre minutes pour obtenir seulement une minute de données, alors qu'avec ce générateur, moins de 30 secondes suffisent pour générer une minute de trafic. Ce gain de temps renforce la crédibilité de l'expérimentation d'attaques très spécialisées en vue du développement de solutions de détection adaptées à ces menaces. Il permet également de répéter plus rapidement des scénarios d'attaques afin de tester d'éventuelles améliorations des heuristiques de détection déjà en production.

Pour aller encore plus loin, ce générateur peut être considérablement amélioré en utilisant des méthodes de type chaîne de Markov pour garantir le respect des machines à état de certains protocoles, ainsi que pour prendre en compte des événements statistiques interdépendants. De plus, envisager la génération du trafic en utilisant un format de type *sFlow* directement injecté sur une interface réseau virtuelle permettrait d'accélérer encore davantage le processus de test. Il serait également intéressant d'assurer, lors de la génération, que le trafic d'attaque et le trafic nominal soient générés en respectant l'ordre chronologique des paquets, éliminant ainsi le besoin de l'outil `mergecap`. Enfin, la possibilité de générer plusieurs heures de trafic en parallèle en fonction des besoins serait un ajout précieux.

Chapitre 5

Conception et implémentation du système de détection

Sommaire

Ce chapitre se consacre à la présentation détaillée de l'architecture d'un système de détection d'attaques DDoS à caractère volumétrique. Cette architecture résulte d'un travail que j'ai mené conjointement avec des ingénieurs en sécurité réseau d'OVHcloud et des chercheurs de l'équipe 2XS. Il convient de souligner que l'objectif de ce chapitre est d'exposer les contraintes techniques ainsi que les fonctionnalités désirées qui ont orienté la conception et la réalisation du système en question, plutôt que de se focaliser sur la description des algorithmes spécifiques de détection.

Initialement, une mise en contexte du système, dans le cadre de son intégration, est donnée afin de cerner la problématique globale. Par la suite, chaque module constitutif du système est expliqué sous un angle fonctionnel. Chaque module est abordé de manière argumentée, démontrant la pertinence des choix opérés dans leurs conception.

Le socle conceptuel de cette architecture repose sur une compréhension des enjeux liés aux attaques DDoS volumétriques et sur une synthèse des pratiques et méthodologies existantes en matières de détection. Les impératifs de performance, de précision et de passage à l'échelle ont guidé la conception de chaque module, conduisant ainsi à l'élaboration d'une structure qui répond à nos besoins. Les contraintes inhérentes à la nature dynamique des attaques DDoS, caractérisées par leur variabilité et leur sophistication croissante, ont dû être prises en compte dans la conception du système.

En conclusion de ce chapitre, il convient de noter qu'il demeure un effort supplémentaire, en vue de consolider le travail ayant conduit au dépôt d'un brevet industriel pour en faire une contribution scientifique à part entière.

5.1 Motivations

La description de cette architecture de détection d'attaques DDoS volumétriques représente la matérialisation de l'aspect industriel de ma thèse. Lorsque j'ai entamé ce travail de recherche, l'objectif d'OVHcloud était de développer un nouveau système visant à remédier aux lacunes du système actuellement en usage.

Ce dernier avait été conçu au début des années 2010 dans le but de remplacer un système de détection propriétaire vendu par une société de services en sécurité informatique.

Enfin, lors de la phase de revue de l'état de l'art, plusieurs techniques de détection d'attaques DDoS ont été identifiées comme pertinentes. Cependant, après une étude plus approfondie des techniques sélectionnées et un travail de développement pour les mettre en situation sur le trafic de production d'OVHcloud, nous n'avons pas réussi à les faire fonctionner conformément aux descriptions disponibles dans les études. Pour certaines, il était impossible de reproduire l'expérience, comme indiqué dans le chapitre 3, car les jeux de données n'étaient pas disponibles. Pour la plupart, des paramètres de calibration étaient manquants dans les études, et nous n'avons pas été en mesure de reproduire les résultats annoncés. Pour les études restantes, telles que [DT15] et [DT19], nous avons pu reproduire les expériences, mais une fois mises en situation sur notre infrastructure, la charge de trafic les rendait inopérantes. C'est avec cette problématique à l'esprit que le système de détection et le découpage en modules pour réduire la charge de travail, décrits ci-dessous, ont été conçus.

5.1.1 Les solutions commerciales

Ce type de système comporte de nombreux avantages, tels que l'expertise de ces entreprises qui se concentrent exclusivement sur la protection contre les cyberattaques. Ces sociétés disposent d'une expertise pointue dans des domaines tels que la détection d'attaques DDoS et disposent souvent des équipes de recherche dédiées pour suivre les nouvelles menaces et les évolutions technologiques. Cette spécialisation permet aux entreprises clientes de bénéficier des dernières avancées en matière de détection sans avoir à effectuer des mises à jour majeures de leurs infrastructures.

De plus, les solutions commerciales sont généralement conçues pour être faciles à déployer et à intégrer dans les environnements informatiques existants, ce qui réduit la complexité et le temps nécessaire pour mettre en place une solution efficace. L'externalisation de la détection d'attaques DDoS permet également aux entreprises de se concentrer sur leur cœur de métier, en laissant des tiers gérer la sécurité de leurs infrastructures, réduisant ainsi la charge de travail liée à la surveillance constante du trafic réseau.

Un argument économique important réside dans la prévisibilité des coûts. Les solutions commerciales sont souvent proposées sous forme d'abonnements ou de contrats de service, ce qui permet de prévoir et de budgétiser les coûts de sécurité de manière plus stable, sans avoir à investir dans le développement et la maintenance de systèmes internes. De plus, ces solutions sont conçues pour être évolutives, ce qui signifie qu'elles peuvent s'adapter aux besoins croissants en termes de trafic et de sécurité à mesure que l'entreprise se développe, bien que cela ait des limites, comme nous le verrons ultérieurement dans cette section.

Enfin, les trois derniers arguments en faveur de ces solutions relèvent davantage de la gestion des risques et des responsabilités. Les entreprises de sécurité informatique offrent souvent un support technique dédié, permettant aux clients d'obtenir de l'aide en cas de problème ou d'attaques en cours, ce qui peut être essentiel pour minimiser les temps d'arrêt et les perturbations. De plus, en cas de faille de sécurité, les entreprises peuvent bénéficier de responsabilités contractuelles claires envers le fournisseur de la solution choisie, ce qui peut contribuer à réduire les risques juridiques et financiers associés aux attaques DDoS. Enfin, certains secteurs, tels que les services financiers ou de santé, sont soumis à des réglementations strictes en matière de sécurité, et les solutions commerciales peuvent aider les entreprises à se conformer à ces réglementations en fournissant des fonctionnalités spécifiques.

5.1.2 Les solutions développées en interne

Cependant, il est important de noter que les arguments que je viens de présenter, qui justifient le choix judicieux d'adopter ce type de protection, ne s'appliquent pas nécessairement aux fournisseurs mais plus lorsque l'infrastructure de l'entreprise est de petite à moyenne taille ou que le cœur de métier est éloigné du domaine informatique.

En effet, les fournisseurs, tels qu'OVHcloud, optent souvent pour des systèmes de détection d'attaques DDoS conçus en interne, plutôt que de recourir à des solutions clés en main proposées par des entreprises spécialisées en sécurité informatique. Les services cloud gèrent des infrastructures massives et diversifiées qui nécessitent des systèmes de détection d'attaques adaptés à leurs besoins spécifiques. En développant leurs propres systèmes, ils peuvent personnaliser les règles de détection pour les adapter de manière optimale à leur environnement et aux types d'attaques auxquelles ils sont les plus susceptibles de faire face.

De plus, étant donné la rapide évolution des attaques DDoS, les fournisseurs doivent être en mesure de réagir rapidement pour protéger leurs clients. En utilisant des solutions internes, ils ont la flexibilité d'apporter des mises à jour et des améliorations en temps réel pour faire face aux nouvelles menaces, même si celles-ci ne concernent que leur propre infrastructure, sans dépendre d'un tiers.

De surcroît, les fournisseurs gèrent un grand nombre de clients, ce qui leur permet de réaliser des économies d'échelle en développant leurs propres solutions de détection d'attaques. Cette économie d'échelle peut se traduire par des coûts réduits pour les clients et une rentabilité accrue pour le fournisseur.

Par ailleurs, les fournisseurs *cloud* ont une connaissance approfondie de leur propre infrastructure, ce qui leur confère un avantage considérable lorsqu'il s'agit de détecter des anomalies ou des comportements suspects. Ils peuvent exploiter cette expertise pour élaborer des systèmes de détection plus efficaces, spécifiquement adaptés à leurs cas d'utilisation.

Il est important de noter que les entreprises de sécurité informatique tierces peuvent avoir accès aux données de trafic de leurs clients afin de détecter les attaques DDoS. Toutefois, les fournisseurs *cloud* attachent généralement une grande importance à la confidentialité des données de leurs clients. En développant leurs propres systèmes de détection, ils ont un meilleur contrôle sur les données sensibles.

Enfin, les fournisseurs sont responsables de la disponibilité et de la performance de leurs services. En ayant un contrôle total sur leurs systèmes de détection, ils peuvent s'assurer qu'ils sont parfaitement adaptés à leurs exigences en matière de performance et de disponibilité.

5.2 Architecture d'un système de détection d'attaques DDoS

Je vais à présent fournir une analyse détaillée de l'architecture du système de détection des attaques volumétriques. Il est essentiel de noter que le cadre opérationnel de ce système demeure identique à celui exposé dans les chapitres précédents, à savoir l'infrastructure d'OVHcloud. La Figure 5.1 représente de manière schématique l'ensemble de cette infrastructure, au sein de laquelle notre système évolue.

À gauche de cette illustration, nous observons divers dispositifs utilisateurs. Ces dispositifs peuvent aussi bien correspondre à nos clients directs qu'à des clients affiliés à d'autres entités. Tous ces éléments sont interconnectés via Internet. Cette interconnexion est placée au centre de l'illustration, et ce sont généralement les fournisseurs d'accès Internet (FAI) qui permettent aux utilisateurs d'accéder à notre infrastructure via divers types de connexions.

Lorsqu'un utilisateur se connecte à l'un des services que nous hébergeons, son FAI est responsable de l'acheminement du trafic réseau vers l'un de nos points de présence (PoP) répartis à travers le monde. Une fois que le trafic du client atteint le PoP géographiquement le plus proche, il pénètre dans notre infrastructure, comme illustré à droite de la Figure 5.1.

Il est important de noter que la Figure 5.1 a été élaborée dans le but de faciliter la compréhension de ce mécanisme, sans chercher à la rendre exhaustive.

Dans la pratique, aucune restriction n'est imposée quant à l'origine du trafic envoyé vers notre infrastructure dans le but de solliciter un de nos services. C'est dans ce contexte que des appareils malveillants, précédemment décrits dans ce manuscrit, peuvent émettre du trafic qualifié de malveillant. Par exemple, un appareil malveillant individuel peut être à l'origine d'une attaque DoS, tandis qu'un groupe d'appareils malveillants, tels que des botnets, peut être responsable d'une attaque DDoS. Ces attaques peuvent générer plusieurs giga — voire téra — bits de trafic dans le but de perturber un service ou notre infrastructure. Ce comportement peut entraîner des perturbations temporaires, voire permanentes, au niveau d'un service ou de notre infrastructure.

C'est dans ce contexte qu'en collaboration avec des ingénieurs d'OVHcloud et des membres de l'équipe 2XS, je développe l'architecture du système de détection que je présente ici.

Ce système, tel que mis en évidence dans les Figures 5.2 et 5.3, s'emploie à vérifier la légitimité du trafic en utilisant différentes stratégies réparties au sein de plusieurs modules, permettant ainsi de déterminer la légitimité du trafic de manière progressive et rigoureuse. Le fonctionnement du système de détection se déroule de la manière suivante : le trafic généré par les clients à l'extérieur de l'infrastructure contient des

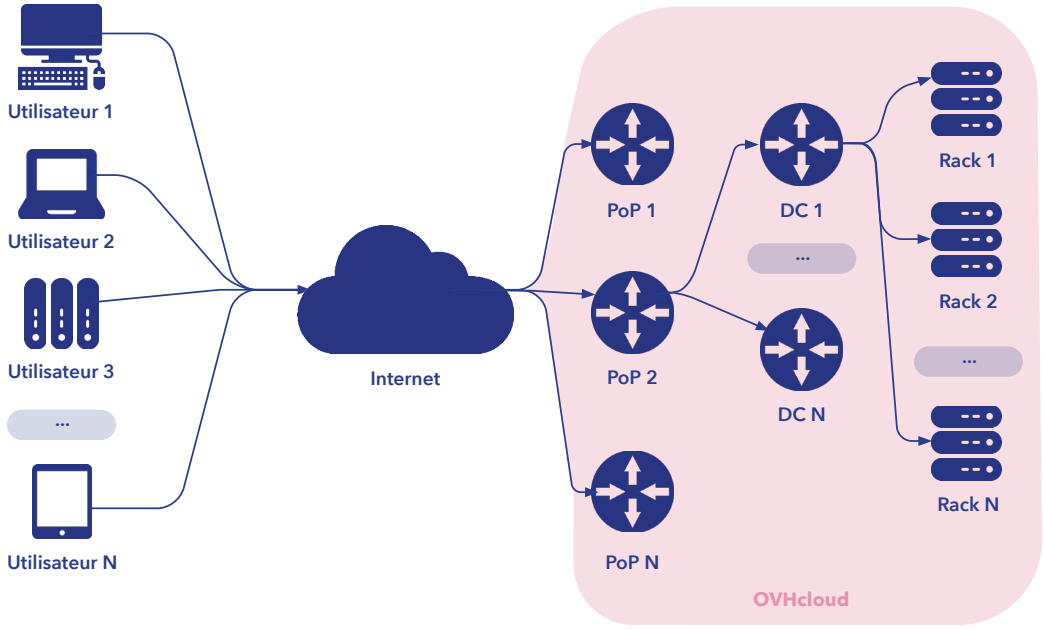

FIGURE 5.1 – Vue simplifiée des échanges entres des appareils et l'infrastructure OVHcloud via Interne

informations cruciales qui sont exploitées pour évaluer la probabilité que le trafic observé soit malveillant. Ces données sont extraites afin d'être mises à disposition des différents modules qui composent le système. Ces informations incluent le couple d'adresses IP source et destination, le couple de ports source et destination, ainsi que le protocole utilisé, qui sont les seules informations présentes dans les différents formats d'exportation de données pris en charge par nos équipements réseau. Toutes ces données sont ensuite transmises aux modules d'analyse.

L'ensemble de ce processus permet au système de détection de procéder à une évaluation progressive et approfondie du trafic entrant. Chaque module d'analyse utilise ces données pour appliquer une série de règles et d'algorithmes qui contribuent à déterminer la nature du trafic : malveillant ou légitime.

Chaque étape du processus d'analyse joue un rôle clé dans la détection d'activités suspectes. Le système fonctionne de manière itérative, en prenant en compte diverses caractéristiques du trafic pour affiner sa conclusion sur la légitimité ou la malveillance. Cette méthodologie en plusieurs étapes garantit une évaluation approfondie et fiable du trafic, ce qui permet de minimiser les fausses alertes tout en identifiant efficacement les menaces potentielles.

Ainsi, grâce à cette approche stratégique et progressive, le système de détection s'assure que seul le trafic légitime est autorisé à accéder à notre infrastructure, tout

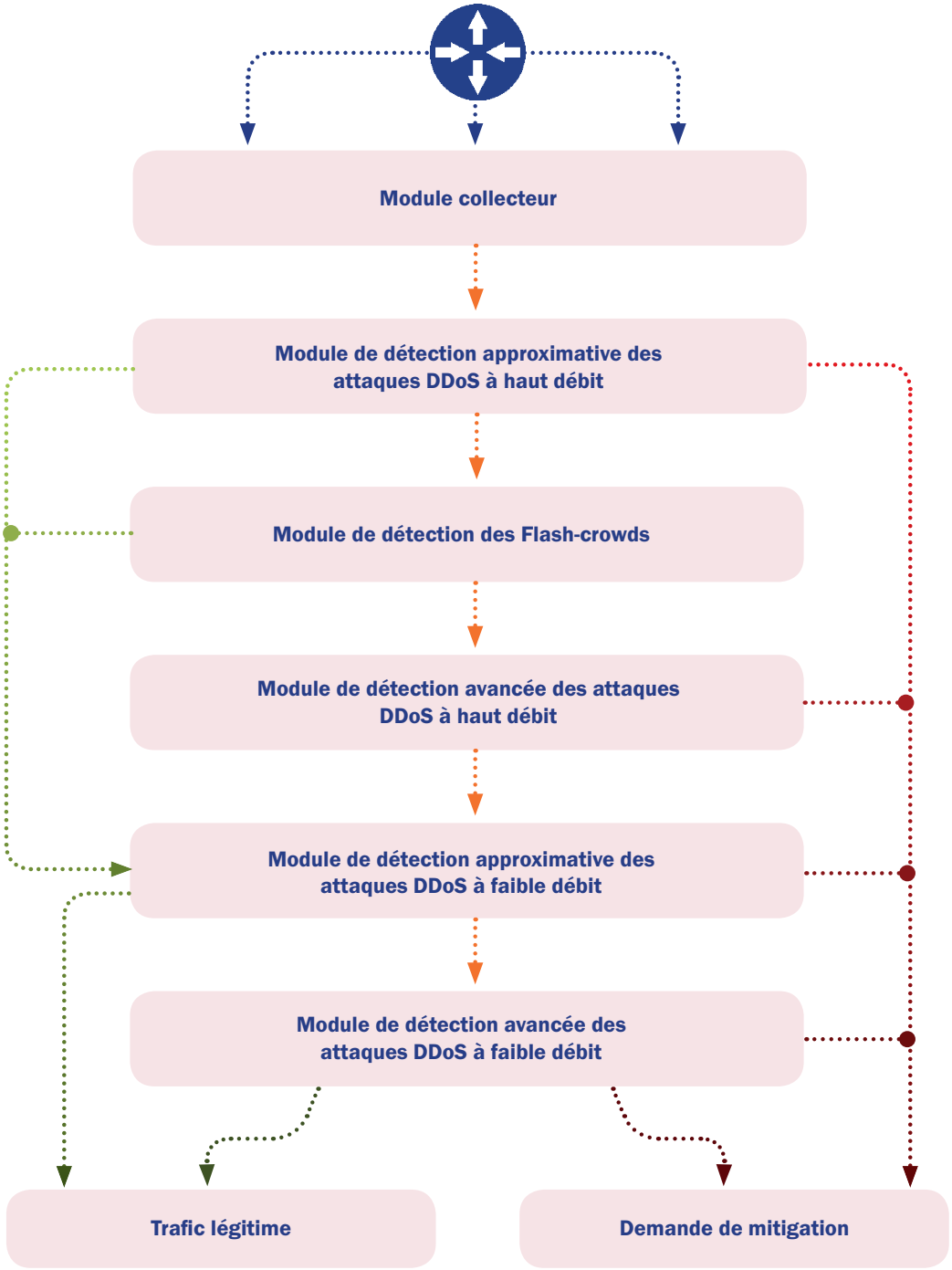

FIGURE 5.2 – Description du fonctionnement du système de détection

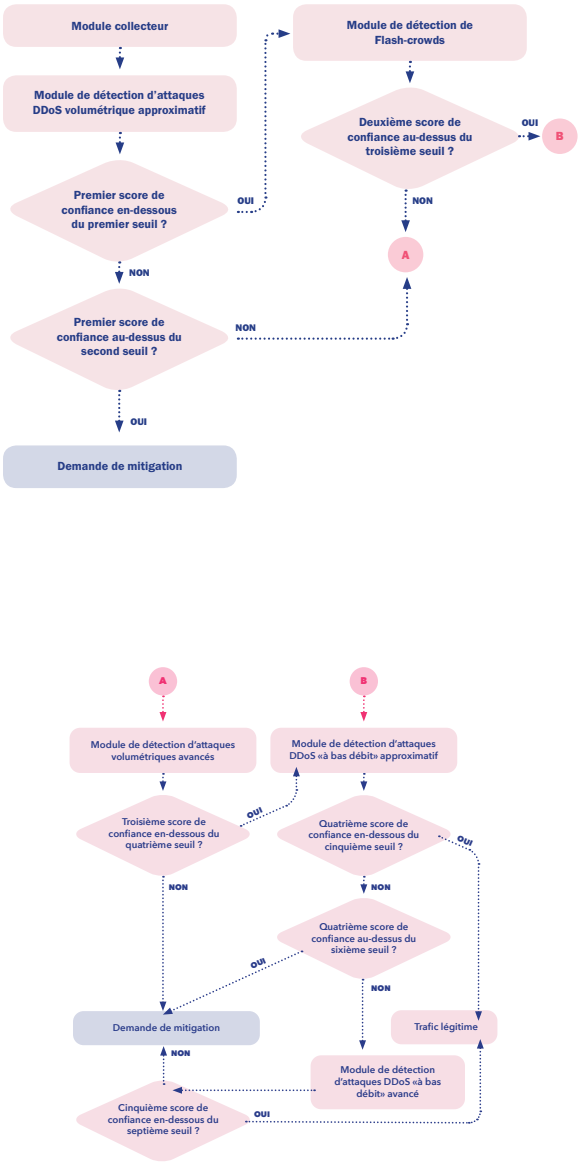

FIGURE 5.3 – Description du pipeline du système de détection

en identifiant rapidement et efficacement les activités malveillantes susceptibles de compromettre la stabilité et la sécurité de notre service.

Dans la suite de ce chapitre, je vais détailler le fonctionnement des différents modules qui constituent le système de détection d'attaques. Tout d'abord, nous avons

le module « collecteur », qui a pour mission de recueillir les événements issus des équipements réseau dispersés dans l'ensemble de notre infrastructure. Son rôle essentiel est d'extraire les informations pertinentes du trafic, qui seront ensuite mises à la disposition des modules suivants.

Ensuite, intervient le module de « détection approximative des attaques ». Ce module exécute un algorithme de détection capable d'établir une première probabilité que le trafic observé soit malveillant parmi l'ensemble du flux entrant dans notre infrastructure. Il constitue une première ligne de défense pour évaluer rapidement la nature du trafic.

Après cela, nous avons le module de « détection de flash-crowds », dont la fonction est de déterminer si le trafic identifié par le module précédent correspond à un pic soudain mais légitime de trafic plutôt qu'à une attaque DDoS. Ce module s'attache à différencier les événements normaux des anomalies temporaires.

Ensuite, si le trafic est identifié comme potentiellement illégitime mais avec une faible probabilité par les deux modules précédents, l'exécution du « module de détection avancée d'attaques » est déclenchée. Ce module, comme nous le verrons ultérieurement, utilise des algorithmes plus précis pour prendre une décision concernant la légitimité du trafic, en tenant compte de diverses caractéristiques et contextes.

Poursuivant dans notre système, nous rencontrons le « module de détection d'attaques DDoS toujours volumétriques », mais représentant un débit beaucoup plus faible, rendant ainsi son identification moins évidente pour les modules précédents. Ce module fonctionne de manière similaire au premier module, mais avec des heuristiques spécifiques adaptées à ce scénario de trafic à plus faible débit.

Enfin, nous concluons avec le « module de détection d'attaques DDoS à faible débit avancé ». À l'instar du module à haut débit, il dispose de davantage de temps et de ressources de calcul pour affiner son calcul de confiance. Il s'attache à détecter les attaques DDoS à faible débit, qui sont souvent plus subtiles et difficiles à repérer.

Ces modules opèrent de manière coordonnée pour évaluer, en plusieurs étapes, la nature du trafic entrant. Chacun d'entre eux apporte une contribution essentielle à la détection précoce des attaques tout en minimisant les fausses alertes. Cette approche en cascade garantit une protection robuste de notre infrastructure contre les menaces potentielles.

5.2.1 Module Collecteur

Ce module constitue le point d'entrée initial du système de détection, chargé de récupérer les informations mentionnées précédemment. Son rôle essentiel est d'extraire les données pertinentes à partir des équipements réseau qui envoient des échantillons de trafic observé. Chez OVHcloud, ces échantillons peuvent prendre trois formats distincts.

Tout d'abord, nous avons le format *NetFlow*, qui est le format historique que nous utilisons. Ce format présente les informations sous forme de résumé de flux, comportant plusieurs champs qui sont mis à jour jusqu'à la détection de la terminaison du flux.

Le deuxième format est *sFlow*, qui, dans la configuration spécifique de notre infrastructure, agit comme un échantillonneur aléatoire. En d'autres termes, selon un taux d'échantillonnage prédéfini, par exemple, 1 pour 2 000, les équipements prennent un paquet dans le flux qu'ils observent, un paquet sur 2 000, et envoient l'en-tête du paquet à notre « module collecteur ».

Enfin, le dernier format est le *kFlow*, un format que nous avons conçu en interne. Il représente un version personnalisé du *NetFlow* et est utilisé pour échantillonner le trafic observé directement au niveau des interfaces de nos hyperviseurs, qui hébergent les différentes machines virtuelles de nos clients. Nous pouvons donc distinguer deux familles de formats : *NetFlow* et *kFlow* d'un côté, et *sFlow* de l'autre, chacune ayant des comportements différents. C'est pourquoi, au sein de ce « module collecteur », nous disposons de deux sous-modules spécialisés pour extraire les informations de ces deux familles de formats.

En effet, la nature même du *NetFlow*, qui agrège les informations échantillonnées sur la durée complète d'un flux, engendre une certaine latence et une perte d'informations due à cette agrégation. En revanche, *sFlow* envoie directement l'en-tête du paquet qui a été tiré au sort, fournissant ainsi des informations « brutes » dès leur capture.

À l'heure actuelle, nous n'avons pas encore trouvé de solution pour harmoniser ces deux familles de formats, ce qui nécessite la mise en place de deux sous-modules distincts. Cette complexité additionnelle impacte la maintenance du « module collecteur » et a également des répercussions sur ses performances, car il doit gérer la répartition des données vers les sous-modules en fonction de leur nature, au lieu de se concentrer uniquement sur l'extraction des données utiles aux modules suivants. Cette problématique est actuellement au centre de nos préoccupations en vue d'améliorer l'efficacité et la cohérence de notre système de détection.

5.2.2 Module de détection d'attaques DDoS volumétriques approximatif

Le module de détection approximative des attaques DDoS volumétriques assume la tâche d'établir une première probabilité quant à l'inclusion du trafic observé dans une campagne d'attaque. Il reçoit en entrée les informations extraites du module précédent et, par exemple, en utilisant un algorithme de détection basé sur des seuils, il évalue si les valeurs calculées pour le trafic observé présentent des variations significatives

par rapport au modèle que nous avons établi du trafic nominal d'un client. Dans ce contexte, l'utilisation de tels algorithmes nécessite la définition de valeurs de seuil pour le trafic considéré comme acceptable. La détection d'une attaque intervient lorsque les valeurs établies sont dépassées.

L'avantage de ces algorithmes réside dans leur relative simplicité de mise en œuvre et leur faible coût en termes de ressources, bien qu'ils puissent générer des fausses alertes ou manquer la détection d'attaques à faible débit. Cependant, ces problématiques sont adressées par les modules que nous examinerons plus en détail dans ce chapitre.

Une autre catégorie d'algorithmes qui pourrait être pertinente à intégrer dans ce module est celle des algorithmes d'analyse de la distribution des paquets. Ces types d'algorithmes analysent la distribution des paquets pour repérer des anomalies, telles qu'une distribution de paquets soudainement asymétrique. Toutefois, ils requièrent souvent une grande quantité de données pour effectuer une analyse statistique fiable et sont également sensibles au « bruit de fond ».

Une évolution intéressante de ce type d'algorithmes que je peux envisager pour ce module serait leur capacité à vérifier la cohérence des séquences de paquets. Cependant, il convient de noter qu'étant donné que nos données d'entrée sont basées sur de l'échantillonnage, ces algorithmes ne puissent pas observer l'ensemble des paquets composant le trafic. De plus, ces algorithmes peuvent nécessiter des ressources mémoire significatives pour stocker les séquences de paquets à notre échelle.

Lorsque le premier score de confiance dépasse un seuil établi, cela signifie que le trafic observé est potentiellement lié à une campagne d'attaque. Pour illustrer cela, reprenons l'exemple où le module utilise un algorithme basé sur des seuils. Dans cette situation, la probabilité d'une attaque est considérée comme élevée, et le module déclenche une notification au système de mitigation pour qu'il procède à la « purge » du trafic malveillant. C'est une mesure proactive visant à neutraliser l'attaque en cours.

En revanche, le deuxième cas d'intérêt se produit lorsque le seuil calculé se situe entre les deux seuils qui permettent soit de conclure qu'il n'y a très probablement pas d'attaque, soit qu'il y a très probablement une attaque en cours. Dans cette situation, le module transmet ses conclusions au « module de détection de *flash-crowds* ». Comme nous le verrons, ce module est conçu pour identifier les pics soudains de trafic liés à des événements légitimes.

5.2.3 Module de détection d'événements de type Flash-crowds

Le rôle du « module de détection de *flash-crowds* » est de distinguer les pics de trafic normaux, tels que ceux provoqués par une forte demande pour un service légitime, des attaques DDoS. Ces pics soudains de trafic peuvent être causés par des

événements populaires, des promotions en ligne, ou d'autres facteurs légitimes qui attirent un grand nombre d'utilisateurs vers un service ou une plateforme. Ce module peut être réalisé en analysant es valeurs d'entropie de l'adresse source et d'entropie des clusters de trafic afin d'obtenir une détection précise des attaques [GB18].

Ce module joue un rôle crucial dans la distinction entre un événement de type *flash-crowd*, caractérisé par un afflux soudain de trafic légitime, et une attaque DDoS. Il évalue cette distinction en calculant un score de confiance basé sur les informations fournies par le « module collecteur » et celles du « module de détection approximative ». Lorsque ce score dépasse un seuil prédéfini, le trafic observé est considéré comme potentiellement lié à une campagne d'attaque, et il transmet ses conclusions au « module de détection avancée » pour une analyse plus approfondie, étant donné qu'il subsiste un doute.

Cependant, lorsque le deuxième score de confiance est associé à un événement de type *flash-crowd*, le module effectue une vérification supplémentaire en sollicitant l'intervention du « module de détection approximative à faible débit ». Cette étape vise à confirmer qu'il ne s'agit pas d'un faux positif. Ainsi, ce processus de double vérification garantit que les actions entreprises en réponse au trafic observé sont adaptées à la nature réelle de la situation, renforçant ainsi la précision de notre système de détection.

5.2.4 Module de détection avancé d'attaques DDoS volumétriques

Ce module, opérant en se basant sur les conclusions du « module de détection des *flash-crowds* », joue un rôle crucial en évaluant la probabilité que le trafic soit associé à une campagne d'attaque. Il se distingue du module approximatif par plusieurs aspects. Comme il ne traite qu'un sous-ensemble du trafic total entrant dans notre infrastructure, il utilise des fenêtres d'acquisition plus longues, ce qui signifie qu'il attend de recevoir plusieurs entrées pour un même flux de trafic de la part des modules précédents. Cette approche lui permet de disposer de davantage d'échantillons statistiques pour calculer son score de confiance, même si cela rallonge le délai pour déclencher une alerte. Cependant, cette méthodologie permet d'obtenir une évaluation plus précise, notamment dans les situations ambiguës. Ce module peut être mis en œuvre en utilisant une méthode de détection d'anomalies basée sur l'entropie de Tsallis [Ana12] et l'exposant de Lyapunov [Din06]. En comparent l'entropie entre les adresses IP source et les adresses IP de destination, en analysant le taux de séparation des exposants [MC13].

On peut considérer le module approximatif comme une étape préliminaire de filtrage pour ce module de détection avancée. Le module approximatif est conçu pour

gérer des volumes massifs de trafic, de l'ordre des térabits, tandis que le module avancé se concentre sur des volumes moins importants, de l'ordre des gigabits. Cette distinction réduit la charge de travail, ce qui libère des ressources et du temps pour l'exécution d'algorithmes de détection plus complexes, demandant davantage de ressources et de temps.

En ce qui concerne le déclenchement du module de mitigation, lorsque le score de confiance calculé est inférieur à un seuil préétabli, une alerte est générée. Cependant, si le score se trouve dans une « zone grise », afin d'éviter les faux positifs, le « module de détection approximatif à faible débit » est notifié. Cette approche réfléchie permet de prendre des mesures appropriées en réponse au trafic observé, garantissant ainsi une meilleure gestion des menaces tout en minimisant les alertes injustifiées.

5.2.5 Module de détection approximatif d'attaques à faible volumétrie

Le module décrit ici prend en entrée les informations provenant des trois modules de détection précédents. Sa principale mission consiste à attribuer un niveau de confiance supplémentaire quant à la probabilité que le trafic observé fasse partie ou non d'une campagne d'attaque.

Son fonctionnement est similaire au « module de détection d'attaques volumétriques approximatif », avec l'exception qu'il se concentre sur la détection d'attaques à plus faible débit. Par conséquent, il utilise des heuristiques différentes, comme les quatre mesures d'entropie de l'information [AF86] : entropie de Hartley, entropie de Shannon, entropie de Renyi++ et entropie généralisée de Renyi++ [BBK15] ou en utilisant une méthode détection est basée sur l'analyse du réseau concernant sa similarité avec lui-même, qui est définie à l'aide du coefficient de Hurst et des caractéristiques propres aux botnets [Lys+20]. La procédure de mise en œuvre ou de non-mise en œuvre du service de mitigation est la même que pour les autres modules. Il est essentiel de souligner que les conclusions tirées des modules de détection d'attaques volumétriques s'appliquent également aux modules de détection d'attaques à faible volumétrie.

Il convient de noter que, bien que nous évoquions ici une « faible volumétrie ou débit », nous opérons toujours dans le cadre d'un fournisseur de services cloud, où ce qui est considéré comme faible peut avoir une signification importante pour des infrastructures plus modestes.

5.2.6 Module de détection avancé d'attaques DDoS à faible volumétrie

Ce module constitue la dernière composante du système de détection, fonctionnant sur les mêmes bases que son homologue dédié aux attaques volumétriques. Cependant, contrairement au module précédent, il met en place des heuristiques, comme celle basées sur la taille attendue des paquets [Zho+17] spécifiques pour la détection d'attaques sur des débits plus modestes, en considérant le point de vue d'un fournisseur de services *cloud*.

Ce dernier module, en tant que phase finale de la détection, joue un rôle crucial. Si le trafic observé n'est toujours pas identifié comme malveillant, il sera considéré comme légitime seulement si le score de confiance calculé est inférieur au seuil prédéfini. Il est important de noter que, même lorsque nous évoquons une « faible volumétrie ou débit » dans ce contexte, nous opérons toujours dans le cadre d'un fournisseur de services *cloud*.

5.2.7 Flux d'éxecution du pipeline de détection

La Figure 5.3 représente le schéma de flux du système de détection. Il commence par l'extraction des données essentielles à la détection des attaques DDoS, telles que les paires d'adresses IP, les ports et le protocole de transport. Ces données sont ensuite transmises au module de détection approximatif des attaques volumétriques. Ce module, en utilisant les informations reçues de l'outil d'extraction de données, établit un premier score de confiance quant à la possibilité que le trafic observé fasse partie d'une campagne d'attaque.

Si le score de confiance se situe dans une « zone grise », le module de détection des *flash-crowds* est activé. En revanche, si le score indique clairement qu'une attaque est en cours, le module de mitigation est notifié. Si le trafic n'est pas associé à un *flash-crowds* par le module prévu à cet effet, alors il est dirigé vers le module de détection avancée des attaques DDoS volumétriques. Ce module, utilisant des ressources de calcul supplémentaires et une période de collecte de données plus longue, établit un nouveau score de confiance.

Si ce score de confiance demeure dans une zone incertaine, le module de détection approximative à faible débit est déclenché. Conformément au fonctionnement de son homologue volumétrique, il établit un score de confiance. Enfin, en dernier recours, le module de détection avancée des attaques à faible débit intervient pour décider s'il faut ou non déclencher une action de mitigation, fonctionnant de manière analogue au module de détection avancée volumétrique.

5.3 Conclusion

En résumé, notre objectif principal était de présenter les contraintes techniques ainsi que les fonctionnalités qui ont guidé la conception de ce système de détection. Nous avons ainsi introduit un système de détection modulaire conçu pour gérer le flux massif de trafic que l'infrastructure d'OVHcloud doit absorber. L'utilisation de modules de détection approximatifs, bien qu'elle puisse occasionnellement entraîner des faux positifs, s'avère indispensable pour faire face à cette charge de travail considérable.

Pour atténuer le risque de faux positifs, nous avons mis en place des modules dits « avancés » qui sont activés sur demande, lorsque le score de confiance se situe dans une « zone grise » et permet de déterminer si le trafic doit être associé à une campagne d'attaque ou non. Cependant, il est important de noter que notre travail sur cette contribution n'est pas encore achevé. Des étapes cruciales restent à franchir, notamment la validation empirique et la réalisation de tests en conditions réelles.

À l'heure où je rédige ce manuscrit, ces deux aspects sont toujours en cours de développement. Une fois que nous aurons recueilli les résultats de ces phases, nous pourrons comparer le système que nous avons décrit ici avec les systèmes de pointe existants. Cette comparaison nous permettra de solliciter les avis de nos pairs afin de renforcer la crédibilité de nos résultats et d'améliorer encore davantage notre solution.

Chapitre 6

Conclusion

Sommaire

Au cours de cette thèse, nous avons entrepris une étude en trois phases visant à approfondir notre compréhension de l'impact des attaques DDoS volumétriques sur les infrastructures des fournisseurs de services *cloud*, à développer une preuve de concept d'un générateur de trafic capable de reproduire avec précision les caractéristiques statistiques du trafic normal et des attaques spécifiques à ces environnements, et enfin, à concevoir une architecture pour un système de détection des attaques DDoS volumétriques opérationnel au sein des infrastructures des fournisseurs.

Ces recherches ont été menées dans le cadre d'une collaboration entre le laboratoire CRIStAL, plus particulièrement l'équipe 2XS, et OVHcloud. Nos travaux se sont concentrés sur deux axes majeurs. Dans un premier temps, nous avons approfondi la caractérisation des attaques au sein des environnements *cloud*, cherchant à saisir la complexité de ces menaces et leur impact spécifique dans ces contextes hautement dynamiques.

Ensuite, nous avons élaboré une preuve de concept d'un générateur de trafic, capable de reproduire fidèlement les modèles de trafic observés dans les opérations normales et d'attaques au sein des infrastructures *cloud*. Ce générateur constitue les prémices d'un outil pour le développement et le test de solutions de détection d'attaques DDoS, tout en favorisant la reproductibilité des travaux de recherche dans ce domaine crucial pour les fournisseurs lorsqu'il sera finalisé et disponible sous licence libre.

Enfin, le deuxième axe a été la conception d'une architecture avancée pour un

système de détection des attaques DDoS volumétriques, spécialement adaptée aux infrastructures des fournisseurs, à l'image d'OVHcloud.

6.1 Limitations

Plusieurs facteurs limitants doivent être pris en compte dans ces travaux. Ils ont été réalisés sur une infrastructure de production, utilisée en continu par des millions de clients, rendant impossible toute perturbation du bon fonctionnement pour mener des expériences. En outre, en raison des volumes de données considérables, dépassant 7 Tbps de trafic brut, même en enregistrant tous les paquets réseau à l'entrée du centre de données le moins sollicité, ce volume dépasse la vitesse d'écriture des SSD les plus performants actuellement disponibles. Ainsi, nous avons dû recourir aux technologies d'échantillonnage mises à disposition par les équipementiers, à savoir NetFlow et sFlow, pour obtenir des échantillons dans le cadre de nos expériences. Toutefois, le volume de trafic représenté par ces échantillons, soit plusieurs centaines de millions d'échantillons réseau remontés par les différents équipements de l'infrastructure, nous a obligés à construire une infrastructure de collecte complexe nécessitant un travail d'ingénierie conséquent.

Un autre aspect crucial à prendre en compte est la protection de la vie privée des clients, un engagement essentiel d'OVHcloud conforme à la réglementation, notamment le RGPD. Ceci nous empêche de rendre publique les données sur lesquelles nous avons effectué nos expérimentations, nous obligeant à élaborer des stratégies pour permettre à la communauté scientifique de reproduire nos expériences.

Par ailleurs, bien que le chapitre 5 aborde la question des algorithmes de détection, même si ces travaux ne proposent pas de nouveaux algorithmes, il évoque la réutilisation d'algorithmes déjà en production, principalement basés sur des seuils. Nous ne pouvons pas divulguer les détails sur la manière dont nous utilisons certains seuils pour décider si le trafic fait partie d'une campagne d'attaques. Contrairement au domaine de la cryptographie, où il est possible de rendre publics les algorithmes car la robustesse de la clé repose sur les paramètres choisis pour la générer, la divulgation des seuils permettrait aux attaquants de contourner nos protections et de réussir à attaquer nos clients.

Enfin, les techniques de détection utilisant le machine learning ne sont pas étudiées dans ce document. À l'échelle d'un fournisseur tel qu'OVHcloud, des questions restent en suspens pour l'utilisation de ces méthodes. La construction d'un jeu de données d'apprentissage pose des défis en raison des volumes de données par seconde, rendant le stockage de ces données difficilement envisageable. Pour construire un jeu de données représentatif, une période d'acquisition d'une semaine pourrait sembler raisonnable pour prendre en compte l'ensemble des effets saisonniers présents dans

le trafic observé. Cependant, le coût et l'infrastructure nécessaires pour construire ce jeu de données constituent un défi à approfondir ultérieurement. De plus, étant donné que la mise sous mitigation du trafic d'un client pourrait entraîner des perturbations, il est nécessaire de garantir l'explicabilité de la prise de décision, une exigence qui mérite une exploration plus approfondie dans le contexte du machine learning.

6.2 Résumé des contributions

Dans le chapitre 3 de mon travail de recherche, j'ai approfondi la question cruciale de la détection des attaques DDoS volumétriques au sein des infrastructures des fournisseurs de services *cloud*. À travers l'examen de différentes métriques traditionnellement utilisées dans la littérature pour repérer ces attaques, j'ai pu démontrer leur pertinence lorsqu'elles sont appliquées à ces ensembles de données spécifiques.

En étudiant le trafic réel provenant de l'infrastructure d'OVHcloud, j'ai également mis en lumière les défis supplémentaires liés à l'échelle et aux signaux faibles que ces attaques représentent dans un tel contexte. Les résultats obtenus ont clairement montré que, bien que les métriques traditionnelles puissent être efficaces pour détecter les attaques DDoS volumétriques ciblant des infrastructures de taille modérée, elles se révèlent pratiquement insuffisantes dans le cadre des environnements typiques des fournisseurs de services *cloud*.

Il est important de souligner que, malgré la prévalence de la problématique de la détection de ces attaques dans la littérature spécialisée, très peu de méthodes sont directement applicables aux fournisseurs de services *cloud*. Cette lacune met en évidence un besoin de développer des solutions spécifiquement adaptées à ces environnements complexes et hautement dynamiques.

Dans le chapitre 4, j'ai abordé la mise en œuvre d'une preuve de concept pour un générateur de trafic qui vise à être statistiquement représentatif du trafic observé au sein des infrastructures de fournisseurs de services *cloud* de grande envergure. Bien que la littérature spécialisée insiste sur l'utilisation de jeux de données provenant de captures réelles de trafic de production, non anonymisés, comme la meilleure option, il est important de noter que la fourniture de tels ensembles de données à la communauté scientifique s'avère impossible, et leur mise en œuvre en interne pour des acteurs majeurs tels qu'OVHcloud est très complexe.

C'est pourquoi j'ai choisi d'explorer l'option d'un générateur de trafic, car il représente, à mon sens, un compromis acceptable entre des données réelles et des données synthétiques. En effet, en veillant à ce que les jeux de données générés reproduisent statistiquement les caractéristiques du trafic observé sur l'infrastructure du fournisseur que nous souhaitons analyser, nous pouvons ainsi effectuer des essais de nouvelles

heuristiques et des tests d'amélioration des heuristiques déjà déployées en production.

L'utilisation de ce générateur présente également l'avantage de préserver la confidentialité, car les données fictives sont employées, éliminant ainsi les risques liés à l'exposition de données sensibles des clients lors des tests de sécurité. Cette approche assure la conformité aux réglementations en matière de protection des données, renforçant ainsi la confiance des clients dans les services *cloud*.

Enfin, le générateur de trafic offrant flexibilité, personnalisation et scalabilité. Ils permettent aux fournisseurs de services de mieux anticiper leurs besoins en capacité, d'expérimenter différentes stratégies de défense etde s'adapter rapidement aux évolutions du paysage des menaces. Cette approche globale contribue à maintenir la disponibilité des services *cloud*, à protéger la confiance des clients et à garantir un environnement plus sûr et résilient face aux attaques DDoS.

Pour conclure, le chapitre 5 dévoile en détail l'architecture du système de détection qui sera bientôt déployée chez OVHcloud. Cette architecture découle d'une réflexion approfondie menée en collaboration avec les deux parties prenantes de ma thèse. Elle est destinée à remplacer l'actuel système de détection des attaques DDoS en production chez OVHcloud.

Les premières preuves de concept que nous avons réalisées indiquent que nous sommes désormais en mesure de gérer la charge générée par l'ensemble des équipements réseau que nous utilisons pour surveiller le trafic transitant par notre infrastructure, le tout sur une seule machine physique puissamment dimensionnée. Cela représente une avancée significative par rapport à l'utilisation de plusieurs dizaines de machines physiques très puissantes, comme cela était nécessaire jusqu'à présent. Cette optimisation de l'infrastructure démontre notre capacité à relever les défis de performance posés par la détection des attaques DDoS à grande échelle.

Il est important de noter que, bien que les heuristiques utilisées au cours des phases de test soient qualifiées de « simples », elles s'avèrent prometteuses et nous mettent sur la bonne voie pour la construction d'un tout nouveau système de détection. Ce système est en passe de répondre pleinement aux objectifs initiaux que nous nous étions fixés au début de ma thèse, à savoir une détection efficace et évolutive des attaques DDoS au sein de l'environnement complexe d'OVHcloud. Cette avancée marque une étape importante dans l'amélioration de la sécurité et de la résilience de nos services face aux menaces persistantes.

6.3 Perspectives de recherche à long terme

Une perspective de recherche réside dans l'amélioration du modèle statistique existants, ce qui implique une évolution multidimensionnelle. Tout d'abord, cela englobe

l'intégration de nouvelles sources de données, allant au-delà des paramètres classiques, pour une représentation plus complète du trafic réseau. Cette amélioration continue est importante pour garder une longueur d'avance sur les attaquants, car ces derniers innovent constamment en développant de nouvelles tactiques et en exploitant de nouvelles vulnérabilités. Pour relever ce défi, les chercheurs peuvent s'engager dans l'exploration de modèles comportementaux plus avancés.

Ces modèles prennent en compte les évolutions temporelles du trafic, les variations saisonnières et les comportements spécifiques à chaque application ou service au sein de l'infrastructure *cloud*. Ces nuances dans la modélisation du comportement du trafic permettent une détection plus précise, en identifiant les anomalies subtiles qui pourraient indiquer une attaque DDoS. Par conséquent, en réduisant les fausses alertes, ces modèles plus raffinés améliorent considérablement l'efficacité de la détection, tout en minimisant les perturbations inutiles pour les opérations normales du fournisseur. Ce niveau de sophistication dans la modélisation du trafic réseau est essentiel pour anticiper les attaques émergentes et les défendre de manière proactive.

Un autre axe de recherche qui pourrait être développé est le développement du générateur de trafic adaptés à une grande échelle avec par exemple, l'intégration de comportements liés à la sécurité dans les générateurs de trafic représente un pas significatif vers des simulations plus réalistes. En incluant des mécanismes de chiffrement, d'authentification et de gestion des identités, ces générateurs peuvent simuler des attaques et des contre-mesures de manière authentique. Cela permet aux chercheurs d'évaluer la robustesse des infrastructures *cloud* face à des scénarios de sécurité complexes, renforçant ainsi la préparation aux cybermenaces. L'utilisation de chaînes de Markov pour générer du trafic réaliste est une approche puissante. Ces chaînes permettent de modéliser avec précision les transitions entre différents états de trafic, qu'il s'agisse de variations de bande passante, de flux de données ou de types d'applications. Cette modélisation fine offre la possibilité de créer des profils de trafic qui respectent pleinement les caractéristiques statistiques d'un fournisseur de services *cloud*, y compris les variations temporelles et saisonnières. De plus, des modèles de chaînes de Markov plus complexes peuvent être explorés pour représenter de manière encore plus précise ces transitions, permettant ainsi une simulation plus réaliste et une évaluation plus précise des performances et de la sécurité.

Mais aussi, la protection de la vie privée et la conformité réglementaire sont des considérations essentielles lors de la collecte et de l'utilisation des données de trafic générées par le générateur. Il est impératif d'examiner en profondeur les implications de ces activités sur la vie privée des utilisateurs et de mettre en place des mécanismes d'anonymisation et de sécurisation des données pour garantir la conformité avec les réglementations en vigueur. Cette démarche renforce la confiance des utilisateurs et garantit que les essais de sécurité respectent les normes légales.

Enfin, l'utilisation du générateur de trafic pour des tests de sécurité exhaustifs, y compris la détection d'attaques DDoS et l'évaluation de la résilience de l'infrastructure cloud, constitue une étape cruciale dans l'amélioration de la sécurité. Ces tests aident à identifier les vulnérabilités, à perfectionner les mécanismes de détection et de réaction, et à assurer la préparation en cas d'attaques réelles. Ils contribuent ainsi à renforcer la sécurité globale des fournisseurs de services cloud, en créant des environnements plus sûrs et plus résilients.

Bibliographie

[ABA17] Hassan Faisal Aldheleai, Mohammad Ubaidullah Bokhari et Abdulsalam Alammari. « Overview of cloud-based learning management system ». In : *International Journal of Computer Applications* 162.11 (2017) (cf. p. 22, 23).

[Abd+22] Mujaheed Abdullahi, Yahia Baashar, Hitham Alhussian, Ayed Alwadain, Norshakirah Aziz, Luiz Fernando Capretz et Said Jadid Abdulkadir. « Detecting cybersecurity attacks in internet of things using artificial intelligence methods : A systematic literature review ». In : *Electronics* 11.2 (2022), p. 198 (cf. p. 41).

[Abh+20] Abhishta Abhishta, Wouter van Heeswijk, Marianne Junger, Lambert JM Nieuwenhuis et Reinoud Joosten. « Why would we get attacked ? An analysis of attacker's aims behind DDoS attacks. » In : *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.* 11.2 (2020), p. 3-22 (cf. p. 33).

[Abh19] Abhishta Abhishta. *The blind man and the elephant : Measuring economic impacts of ddos attacks*. University of Twente, 2019 (cf. p. 14).

[AF86] J Aczel et B Forte. « Generalized entropies and the maximum entropy principle ». In : *Maximum Entropy and Bayesian Methods in Applied Statistics* (1986), p. 95-100 (cf. p. 111).

[AGM16] Mohamed Almorsy, John Grundy et Ingo Müller. « An analysis of the cloud computing security problem ». In : *arXiv preprint arXiv :1609.01107* (2016) (cf. p. 31).

[AH18] Sabah Alzahrani et Liang Hong. « Detection of distributed denial of service (DDoS) attacks using artificial intelligence on cloud ». In : *2018 IEEE World Congress on Services (SERVICES)*. IEEE. 2018, p. 35-36 (cf. p. 41).

[Al-+13] May Al-Roomi, Shaikha Al-Ebrahim, Sabika Buqrais et Imtiaz Ahmad. « Cloud computing pricing models : a survey ». In : *International Journal of Grid and Distributed Computing* 6.5 (2013), p. 93-106 (cf. p. 20).

[Ali+16] Kamal Alieyan, Mohammed M Kadhum, Mohammed Anbar, Shafiq Ul Rehman et Naser KA Alajmi. « An overview of DDoS attacks based on DNS ». In : *2016 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE. 2016, p. 276-280 (cf. p. 35).

[Alk+16] Mouhammd Alkasassbeh, Ghazi Al-Naymat, Ahmad BA Hassanat et Mohammad Almseidin. « Detecting distributed denial of service attacks using data mining techniques ». In : *International Journal of Advanced Computer Science and Applications* 7.1 (2016) (cf. p. 63).

[AM+16] BS Amulya, Minal Moharir et al. « A study of tools to develop a traffic generator for L4–L7 layers ». In : *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*. IEEE. 2016, p. 114-118 (cf. p. 89).

[Ana12] Anastasios Anastasiadis. « Tsallis entropy ». In : *Entropy* 14.2 (2012), p. 174-176 (cf. p. 110).

[And+06] Georgios Androulidakis, Vasilis Chatzigiannakis, Symeon Papavassiliou, Mary Grammatikou et Vasilis Maglaris. « Understanding and evaluating the impact of sampling on anomaly detection techniques ». In : *MILCOM 2006-2006 IEEE Military Communications conference*. IEEE. 2006, p. 1-7 (cf. p. 80).

[Ant+17] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J Alex Halderman, Luca Invernizzi, Michalis Kallitsis et al. « Understanding the mirai botnet ». In : *26th USENIX security symposium (USENIX Security 17)*. 2017, p. 1093-1110 (cf. p. 38).

[AR12] Mohammed Alenezi et Martin J Reed. « Methodologies for detecting DoS/DDoS attacks against network servers ». In : *The Seventh International Conference on Systems and Networks Communications ICSNC*. 2012, p. 92-98 (cf. p. 40).

[BBK15] Monowar H Bhuyan, DK Bhattacharyya et Jugal K Kalita. « An empirical evaluation of information metrics for low-rate and high-rate DDoS attack detection ». In : *Pattern Recognition Letters* 51 (2015), p. 1-7 (cf. p. 111).

[BCL21] Will Bonasera, Md Minhaz Chowdhury et Shadman Latif. « Denial of service : A growing underrated threat ». In : *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. IEEE. 2021, p. 1-6 (cf. p. 14).

[BD18] Shweta M Barhate et MP Dhore. « Hybrid cloud : A solution to cloud interoperability ». In : *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. IEEE. 2018, p. 1242-1247 (cf. p. 24).

[Ben+18] Ahmed Bentajer, Mustapha Hedabou, Karim Abouelmehdi et Said Elfezazi. « CS-IBE : a data confidentiality system in public cloud storage system ». In : *Procedia computer science* 141 (2018), p. 559-564 (cf. p. 23).

[Bha+16] Akashdeep Bhardwaj, GVB Subrahmanyam, Vinay Avasthi et Hanumat G Sastry. « Solutions for DDoS attacks on cloud ». In : *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence)*. IEEE. 2016, p. 163-167 (cf. p. 45).

[BHR19] Luiz André Barroso, Urs Hölzle et Parthasarathy Ranganathan. *The datacenter as a computer : Designing warehouse-scale machines*. Springer Nature, 2019 (cf. p. 30, 32).

[Bil+14] Kashif Bilal, Saif Ur Rehman Malik, Samee U Khan et Albert Y Zomaya. « Trends and challenges in cloud datacenters ». In : *IEEE cloud computing* 1.1 (2014), p. 10-20 (cf. p. 31).

[BJJ10] Sushil Bhardwaj, Leena Jain et Sandeep Jain. « Cloud computing : A study of infrastructure as a service (IAAS) ». In : *International Journal of engineering and information Technology* 2.1 (2010), p. 60-63 (cf. p. 21).

[BK16] Sunny Behal et Krishan Kumar. « Trends in validation of DDoS research ». In : *Procedia Computer Science* 85 (2016), p. 7-15 (cf. p. 62).

[BN20] Maria Bada et Jason RC Nurse. « The social and psychological impact of cyberattacks ». In : *Emerging cyber threats and cognitive vulnerabilities*. Elsevier, 2020, p. 73-92 (cf. p. 14).

[Boi+23] Clément Boin, Tristan Groléat, Xavier Guillaume, Gilles Grimaud et Michaël Hauspie. « Scale matters : a Comparative Study of Datasets for DDoS Attack Detection in CSP Infrastructure ». In : *CloudNet2023*. 2023 (cf. p. 17).

[BR07] Boldizsár Bencsáth et Miklós Aurél Rónai. « Empirical analysis of Denial of Service attack against SMTP servers ». In : *2007 International Symposium on Collaborative Technologies and Systems*. IEEE. 2007, p. 72-79 (cf. p. 36).

[Bre12] Susan W Brenner. *Cybercrime and the law : Challenges, issues, and outcomes*. UPNE, 2012 (cf. p. 38).

[Bro+21] Richard R Brooks, Lu Yu, Ilker Ozcelik, Jon Oakley et Nathan Tusing. « Distributed denial of service (DDoS) : a history ». In : *IEEE Annals of the History of Computing* 44.2 (2021), p. 44-54 (cf. p. 37).

[BS14] Sangeeta Bhattacharya et S Selvakumar. « SSENet-2014 dataset : A dataset for detection of multiconnection attacks ». In : *2014 3rd International Conference on Eco-friendly Computing and Communication Systems*. IEEE. 2014, p. 121-126 (cf. p. 63).

[Cam+18] Carla O Camargo, Elaine R Faria, Bruno B Zarpelão et Rodrigo S Miani. « Qualitative evaluation of denial of service datasets ». In : *Proceedings of the XIV Brazilian Symposium on Information Systems*. 2018, p. 1-8 (cf. p. 62).

[Cat+21] Marta Catillo, Antonio Pecchia, Massimiliano Rak et Umberto Villano. « Demystifying the role of public intrusion datasets : A replication study of DoS network traffic data ». In : *Computers & Security* 108 (2021), p. 102341 (cf. p. 63).

[Cha05] Steven Chamberland. « Point of presence design in Internet protocol networks with performance guarantees ». In : *Computers & operations research* 32.12 (2005), p. 3247-3264 (cf. p. 30).

[Cor+15] Carlos Garcia Cordero, Emmanouil Vasilomanolakis, Nikolay Milanov, Christian Koch, David Hausheer et Max Mühlhäuser. « ID2T : A DIY dataset creation toolkit for intrusion detection systems ». In : *2015 IEEE Conference on Communications and Network Security (CNS)*. IEEE. 2015, p. 739-740 (cf. p. 88).

[CYC13] Johannes K Chiang, Eric H-W Yen et Yen-Hua Chen. « Authentication, authorization and file synchronization in hybrid cloud : On case of Google Docs, Hadoop and Linux local hosts ». In : *2013 International Symposium on Biometrics and Security Technologies*. IEEE. 2013, p. 116-123 (cf. p. 24).

[Dag+07] David Dagon, Guofei Gu, Christopher P Lee et Wenke Lee. « A taxonomy of botnet structures ». In : *Twenty-Third Annual Computer Security Applications Conference (ACSAC 2007)*. IEEE. 2007, p. 325-339 (cf. p. 34).

[Dam+20] Robertas Damasevicius, Algimantas Venckauskas, Sarunas Grigaliunas, Jevgenijus Toldinas, Nerijus Morkevicius, Tautvydas Aleliunas et Paulius Smuikys. « LITNET-2020 : An annotated real-world network flow dataset for network intrusion detection ». In : *Electronics* 9.5 (2020), p. 800 (cf. p. 62, 63, 84).

[DC19] Smita Dange et Madhumita Chatterjee. « IoT botnet : The largest threat to the IoT network ». In : *Data Communication and Networks : Proceedings of GUCON 2019*. Springer, 2019, p. 137-157 (cf. p. 38).

[DD15] Rashmi V Deshmukh et Kailas K Devadkar. « Understanding DDoS attack & its effect in cloud environment ». In : *Procedia Computer Science* 49 (2015), p. 202-210 (cf. p. 36).

[Din06] Jonathan B Dingwell. « Lyapunov exponents ». In : *Wiley encyclopedia of biomedical engineering* (2006) (cf. p. 110).

[Dit99] David Dittrich. *The 'stacheldraht'distributed denial of service attack tool*. 1999 (cf. p. 37).

[DL00] Sven Dietrich et Neil Long. « Analyzing distributed denial of service tools : The shaft case ». In : *14th Systems Administration Conference (LISA 2000)*. 2000 (cf. p. 37).

[DPM21] Marinos Dimolianis, Adam Pavlidis et Vasilis Maglaris. « Signature-based traffic classification and mitigation for DDoS attacks using programmable network data planes ». In : *IEEE Access* 9 (2021), p. 113061-113076 (cf. p. 40).

[DPŞ14] M Nishan Dharmaweera, Rajendran Parthiban et Y Ahmet Şekercioğlu. « Toward a power-efficient backbone network : The state of research ». In : *IEEE Communications Surveys & Tutorials* 17.1 (2014), p. 198-227 (cf. p. 32).

[DR17a] Tinankoria Diaby et Babak Bashari Rad. « Cloud computing : a review of the concepts and deployment models ». In : *International Journal of Information Technology and Computer Science* 9.6 (2017), p. 50-58 (cf. p. 20).

[DR17b] Tinankoria Diaby et Babak Bashari Rad. « Cloud computing : a review of the concepts and deployment models ». In : *International Journal of Information Technology and Computer Science* 9.6 (2017), p. 50-58 (cf. p. 22).

[DT15] Jisa David et Ciza Thomas. « DDoS attack detection using fast entropy approach on flow-based network traffic ». In : *Procedia Computer Science* 50 (2015), p. 30-36 (cf. p. 100).

[DT19] Jisa David et Ciza Thomas. « Efficient DDoS flood attack detection using dynamic thresholding on flow-based network traffic ». In : *Computers & Security* 82 (2019), p. 284-295 (cf. p. 100).

[Dua+13] Jiaqi Duan, Parwiz Faker, Alexander Fesak et Tim Stuart. « Benefits and drawbacks of cloud-based versus traditional ERP systems ». In : *Proceedings of the 2012-13 course on Advanced Resource Planning* (2013) (cf. p. 23).

[Ell00] John Elliott. « Distributed denial of service attacks and the zombie ant effect ». In : *IT professional* 2.02 (2000), p. 55-57 (cf. p. 37).

[Emm+15] Paul Emmerich, Sebastian Gallenmüller, Daniel Raumer, Florian Wohlfart et Georg Carle. « Moongen : A scriptable high-speed packet generator ». In : *Proceedings of the 2015 Internet Measurement Conference*. 2015, p. 275-287 (cf. p. 89).

[Fei+03] Laura Feinstein, Dan Schnackenberg, Ravindra Balupari et Darrell Kindred. « Statistical approaches to DDoS attack detection and response ». In : *Proceedings DARPA information survivability conference and exposition*. T. 1. IEEE. 2003, p. 303-314 (cf. p. 44).

[Fis+13] Alex Fishman, Mike Rapoport, Evgeny Budilovsky et Izik Eidus. « {HVX} : Virtualizing the Cloud ». In : *5th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 13)*. 2013 (cf. p. 23).

[FJ09] AR Flo et Audun Josang. « Consequences of botnets spreading to mobile devices ». In : *Short-Paper Proceedings of the 14th Nordic Conference on Secure IT Systems (NordSec 2009)*. Citeseer. 2009, p. 37-43 (cf. p. 34).

[Fur10] Borko Furht. « Cloud computing fundamentals ». In : *Handbook of cloud computing* (2010), p. 3-19 (cf. p. 20, 30).

[GB18] Jaideep Gera et Bhanu Prakash Battula. « Detection of spoofed and non-spoofed DDoS attacks and discriminating them from flash crowds ». In : *EURASIP Journal on Information Security* 2018 (2018), p. 1-12 (cf. p. 110).

[Gir+15] Anteneh Girma, Moses Garuba, Jiang Li et Chunmei Liu. « Analysis of DDoS attacks and an introduction of a hybrid statistical model to detect DDoS attacks on cloud computing environment ». In : *2015 12th International Conference on Information Technology-New Generations*. IEEE. 2015, p. 212-217 (cf. p. 44).

[GMB16] Brinton Christopher G., Chiang Mung et Christopher G. Brinton. *The power of networks : six principles that connect our lives*. eng. Princeton : Princeton University Press, 2016. isbn : 978-14-0088-407-0 (cf. p. 27).

[Gog+12] Prasanta Gogoi, Monowar H Bhuyan, Dhruba Kumar Bhattacharyya et Jugal K Kalita. « Packet and flow based network intrusion dataset ». In : *Contemporary Computing : 5th International Conference, IC3 2012, Noida, India, August 6-8, 2012. Proceedings 5*. Springer. 2012, p. 322-334 (cf. p. 63).

[GS14] A Gangwar et S Sahu. « A survey on anomaly and signature based intrusion detection system (IDS) ». In : *International Journal of Engineering Research and Applications* 4.4 (2014) (cf. p. 40).

[GW00] Xianjun Geng et Andrew B Whinston. « Defeating distributed denial of service attacks ». In : *It Professional* 2.4 (2000), p. 36-42 (cf. p. 37).

[Hab+12] M Farhan Habib, Massimo Tornatore, Marc De Leenheer, Ferhat Dikbiyik et Biswanath Mukherjee. « Design of disaster-resilient optical datacenter networks ». In : *Journal of Lightwave Technology* 30.16 (2012), p. 2563-2573 (cf. p. 31).

[Ham19] Siham Hamadah. « Cloud-based disaster recovery and planning models : An overview ». In : *ICIC Express Lett* 13.7 (2019), p. 593-599 (cf. p. 25).

[HBK15] Nazrul Hoque, Dhruba K Bhattacharyya et Jugal K Kalita. « Botnet in DDoS attacks : trends and challenges ». In : *IEEE Communications Surveys & Tutorials* 17.4 (2015), p. 2242-2270 (cf. p. 33, 34).

[HKP13] Bryan Harris, Eli Konikoff et Phillip Petersen. « Breaking the DDoS attack chain ». In : *Institute for Software Research* (2013), p. 1-16 (cf. p. 38).

[Hna+22] Mohamad Hnayno, Ali Chehade, Henryk Klaba, Hadrien Bauduin, Guillaume Polidori et Chadi Maalouf. « Performance analysis of new liquid cooling topology and its impact on data centres ». In : *Applied Thermal Engineering* 213 (2022), p. 118733 (cf. p. 31).

[Hol+08] Thorsten Holz, Moritz Steiner, Frederic Dahl, Ernst W Biersack, Felix C Freiling et al. « Measurements and Mitigation of Peer-to-Peer-based Botnets : A Case Study on Storm Worm. » In : *Leet* 8.1 (2008), p. 1-9 (cf. p. 38).

[Hon+19] Jiangshui Hong, Thomas Dreibholz, Joseph Adam Schenkel et Jiaxi Alessia Hu. « An overview of multi-cloud computing ». In : *Web, Artificial Intelligence and Network Applications : Proceedings of the Workshops of the 33rd International Conference on Advanced Information Networking and Applications (WAINA-2019) 33*. Springer. 2019, p. 1055-1068 (cf. p. 26).

[HS14] Neminath Hubballi et Vinoth Suryanarayanan. « False alarm minimization techniques in signature-based intrusion detection systems : A survey ». In : *Computer Communications* 49 (2014), p. 1-17 (cf. p. 40, 41).

[HSM15] William Hurst, Nathan Shone et Quentin Monnet. « Predicting the effects of DDoS attacks on a network of critical infrastructures ». In : *2015 IEEE International Conference on Computer and Information Technology ; Ubiquitous Computing and Communications ; Dependable, Autonomic and Secure Computing ; Pervasive Intelligence and Computing*. IEEE. 2015, p. 1697-1702 (cf. p. 34).

[HW10a] Paul Hofmann et Dan Woods. « Cloud computing : The limits of public clouds for business applications ». In : *IEEE Internet Computing* 14.6 (2010), p. 90-93 (cf. p. 22).

[HW10b] Paul Hofmann et Dan Woods. « Cloud computing : The limits of public clouds for business applications ». In : *IEEE Internet Computing* 14.6 (2010), p. 90-93 (cf. p. 22).

[Jaz+17] Hossein Hadian Jazi, Hugo Gonzalez, Natalia Stakhanova et Ali A Ghorbani. « Detecting HTTP-based application layer DoS attacks on web servers in the presence of sampling ». In : *Computer Networks* 121 (2017), p. 25-36 (cf. p. 63).

[JD17] Khundrakpam Johnson Singh et Tanmay De. « Mathematical modelling of DDoS attack and detection using correlation ». In : *Journal of cyber security technology* 1.3-4 (2017), p. 175-186 (cf. p. 41).

[JG+11] Wayne Jansen, Tim Grance et al. « Guidelines on security and privacy in public cloud computing ». In : (2011) (cf. p. 23).

[KDH21] Daniel Kopp, Christoph Dietzel et Oliver Hohlfeld. « DDoS never dies ? An IXP perspective on DDoS amplification attacks ». In : *International Conference on Passive and Active Network Measurement*. Springer. 2021, p. 284-301 (cf. p. 33).

[Kes00] Gary C Kessler. « Defenses against distributed denial of service attacks ». In : *SANS Institute* 2002 (2000) (cf. p. 37).

[KES21] Ilhan Firat Kilincer, Fatih Ertam et Abdulkadir Sengur. « Machine learning methods for cyber security intrusion detection : Datasets and comparative study ». In : *Computer Networks* 188 (2021), p. 107840 (cf. p. 63).

[Kha+18] Mutaz HH Khairi, Sharifah HS Ariffin, NM Latiff, AS Abdullah et MK Hassan. « A Review of Anomaly Detection Techniques and Distributed Denial of Service (DDoS) on Software Defined Network (SDN). » In : *Engineering, Technology & Applied Science Research* 8.2 (2018) (cf. p. 44).

[KM15] Burak Kantarci et Hussein T Mouftah. « Resilient design of a cloud system over an optical backbone ». In : *IEEE Network* 29.4 (2015), p. 80-87 (cf. p. 32).

[Kop+19] Daniel Kopp, Matthias Wichtlhuber, Ingmar Poese, Jair Santanna, Oliver Hohlfeld et Christoph Dietzel. « DDoS hide & seek : on the effectiveness of a booter services takedown ». In : *Proceedings of the Internet Measurement Conference*. 2019, p. 65-72 (cf. p. 39).

[KR10] Eric Keller et Jennifer Rexford. « The" Platform as a Service" Model for Networking. » In : *INM/WREN* 10 (2010), p. 95-108 (cf. p. 21).

[Kui+15] Hendrik Kuijs, Christoph Reich, Martin Knahl et Nathan Clarke. « Towards privacy for ambient assisted living in a hybrid cloud environment ». In : *BW-CAR/ SINCOM* (2015), p. 41 (cf. p. 25).

[Lac11] Georg Lackermair. « Hybrid cloud architectures for the online commerce ». In : *Procedia Computer Science* 3 (2011), p. 550-555 (cf. p. 24).

[LC09] Simon Liu et Bruce Cheng. « Cyberattacks : Why, what, who, and how ». In : *IT professional* 11.3 (2009), p. 14-21 (cf. p. 13).

[LH21] Olivier Lourme et Michaël Hauspie. « Toward a realistic Intrusion Detection System dedicated to smart-home environments ». In : *2021 17th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE. 2021, p. 80-85 (cf. p. 38).

[Li+10] Ang Li, Xiaowei Yang, Srikanth Kandula et Ming Zhang. « CloudCmp : comparing public cloud providers ». In : *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. 2010, p. 1-14 (cf. p. 22).

[Liu+09] Jing Liu, Yang Xiao, Kaveh Ghaboosi, Hongmei Deng et Jingyuan Zhang. « Botnet : classification, attacks, detection, tracing, and preventive measures ». In : *EURASIP journal on wireless communications and networking* 2009 (2009), p. 1-11 (cf. p. 33, 35).

[LL05] Lan Li et Gyungho Lee. « DDoS attack detection and wavelets ». In : *Telecommunication Systems* 28 (2005), p. 435-451 (cf. p. 41).

[LMN19] Alma D Lopez, Asha P Mohan et Sukumaran Nair. « Network traffic behavioral analytics for detection of DDoS attacks ». In : *SMU data science review* 2.1 (2019), p. 14 (cf. p. 41).

[Lys+20] Sergii Lysenko, Kira Bobrovnikova, Serhii Matiukh, Ivan Hurman et Oleg Savenko. « Detection of the botnets' low-rate DDoS attacks based on self-similarity ». In : *International Journal of Electrical and Computer Engineering* 10.4 (2020), p. 3651-3659 (cf. p. 111).

[Mac+18] Gabriel Maciá-Fernández, José Camacho, Roberto Magán-Carrión, Pedro García-Teodoro et Roberto Therón. « UGR '16 : A new dataset for the evaluation of cyclostationarity-based network IDSs ». In : *Computers & Security* 73 (2018), p. 411-424 (cf. p. 63, 85).

[Mao+06] Z Morley Mao, Vyas Sekar, Oliver Spatscheck, Jacobus Van Der Merwe et Rangarajan Vasudevan. « Analyzing large DDoS attacks using multiple data sources ». In : *Proceedings of the 2006 SIGCOMM workshop on Large-scale attack defense*. 2006, p. 161-168 (cf. p. 35).

[Mar+08] Athina Markopoulou, Gianluca Iannaccone, Supratik Bhattacharyya, Chen-Nee Chuah, Yashar Ganjali et Christophe Diot. « Characterization of failures in an operational IP backbone network ». In : *IEEE/ACM transactions on networking* 16.4 (2008), p. 749-762 (cf. p. 32).

[MC13] Xinlei Ma et Yonghong Chen. « DDoS detection method based on chaos analysis of network traffic entropy ». In : *IEEE Communications Letters* 18.1 (2013), p. 114-117 (cf. p. 110).

[MDL17] Vincenzo Matta, Mario Di Mauro et Maurizio Longo. « DDoS attacks with randomized traffic innovation : Botnet identification challenges and strategies ». In : *IEEE Transactions on Information Forensics and Security* 12.8 (2017), p. 1844-1859 (cf. p. 35).

[Men97] Belden Menkus. « Understanding the Denial of Service Threat ». In : *EDPACS : The EDP Audit, Control, and Security Newsletter* 24.9 (1997), p. 11-17 (cf. p. 37).

[Mia+15] Rui Miao, Rahul Potharaju, Minlan Yu et Navendu Jain. « The dark menace : Characterizing network-based attacks in the cloud ». In : *Proceedings of the 2015 Internet Measurement Conference*. 2015, p. 169-182 (cf. p. 35).

[MK20] Mohammad Masdari et Hemn Khezri. « A survey and taxonomy of the fuzzy signature-based intrusion detection systems ». In : *Applied Soft Computing* 92 (2020), p. 106301 (cf. p. 40).

[Mou+20] Giovane CM Moura, Cristian Hesselman, Gerald Schaapman, Nick Boerman et Octavia De Weerdt. « Into the DDoS maelstrom : a longitudinal study of a scrubbing service ». In : *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE. 2020, p. 550-558 (cf. p. 45).

[MR04] Jelena Mirkovic et Peter Reiher. « A taxonomy of DDoS attack and DDoS defense mechanisms ». In : *ACM SIGCOMM Computer Communication Review* 34.2 (2004), p. 39-53 (cf. p. 33).

[MS15] Nour Moustafa et Jill Slay. « UNSW-NB15 : a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) ». In : *2015 military communications and information systems conference (MilCIS)*. IEEE. 2015, p. 1-6 (cf. p. 63).

[MS21] Hafsa Mateen et Malik Shahzad. « Factors Effecting Businesses due to Distributed Denial of Service (DDoS) Attack ». In : *2021 International Conference on Innovative Computing (ICIC)*. IEEE. 2021, p. 1-7 (cf. p. 39).

[Naz08] Jose Nazario. « DDoS attack evolution ». In : *Network Security* 2008.7 (2008), p. 7-10 (cf. p. 37).

[New19] Sean Newman. « Surviving ransom driven DDoS extortion campaigns ». In : *Cyber Security : A Peer-Reviewed Journal* 3.1 (2019), p. 37-43 (cf. p. 38).

[NKD22] Yevheniya Nosyk, Maciej Korczyński et Andrzej Duda. « Routing loops as mega amplifiers for dns-based ddos attacks ». In : *International Conference on Passive and Active Network Measurement*. Springer. 2022, p. 629-644 (cf. p. 39).

[Ols05] Robert Olsson. « Pktgen the linux packet generator ». In : *Proceedings of the Linux Symposium, Ottawa, Canada*. T. 2. 2005, p. 11-24 (cf. p. 89).

[Pas+17] Túlio A Pascoal, Yuri G Dantas, Iguatemi E Fonseca et Vivek Nigam. « Slow TCAM exhaustion DDoS attack ». In : *ICT Systems Security and Privacy Protection : 32nd IFIP TC 11 International Conference, SEC 2017, Rome, Italy, May 29-31, 2017, Proceedings 32*. Springer. 2017, p. 17-31 (cf. p. 35).

[Pro18] Danijela D Protić. « Review of KDD Cup '99, NSL-KDD and Kyoto 2006+ datasets ». In : *Vojnotehnički glasnik/Military Technical Courier* 66.3 (2018), p. 580-596 (cf. p. 63).

[PT18] Amit Praseed et P Santhi Thilagam. « DDoS attacks at the application layer : Challenges and research perspectives for safeguarding web applications ». In : *IEEE Communications Surveys & Tutorials* 21.1 (2018), p. 661-685 (cf. p. 36).

[Put+15] Deepak Puthal, Bibhudutta PS Sahoo, Sambit Mishra et Satyabrata Swain. « Cloud computing features, issues, and challenges : a big picture ». In : *2015 International conference on computational intelligence and networks*. IEEE. 2015, p. 116-123 (cf. p. 22).

[Rao+16] B Thirumala Rao et al. « A study on data storage security issues in cloud computing ». In : *Procedia Computer Science* 92 (2016), p. 128-135 (cf. p. 31).

[Rao+20] Yerra Shankar Rao, Ajit Kumar Keshri, Bimal Kumar Mishra et Tarini Charana Panda. « Distributed denial of service attack on targeted resources in a computer network for critical infrastructure : A differential e-epidemic model ». In : *Physica A : Statistical Mechanics and Its Applications* 540 (2020), p. 123240 (cf. p. 39).

[RI15] Lauren Rudman et B Irwin. « Characterization and analysis of NTP amplification based DDoS attacks ». In : *2015 Information Security for South Africa (ISSA)*. IEEE. 2015, p. 1-5 (cf. p. 35).

[Rin+19] Markus Ring, Sarah Wunderlich, Deniz Scheuring, Dieter Landes et Andreas Hotho. « A survey of network-based intrusion detection data sets ». In : *Computers & Security* 86 (2019), p. 147-167 (cf. p. 62).

[RMD22] Soumya Ray, Kamta Nath Mishra et Sandip Dutta. « Detection and prevention of DDoS attacks on M-healthcare sensitive data : a novel approach ». In : *International Journal of Information Technology* 14.3 (2022), p. 1333-1341 (cf. p. 39).

[RMG13] Rafael A Rodriguez-Gomez, Gabriel Macia-Fernandez et Pedro Garcia-Teodoro. « Survey and taxonomy of botnet research through life-cycle ». In : *ACM Computing Surveys (CSUR)* 45.4 (2013), p. 1-33 (cf. p. 34).

[Rod] Leonardo Rodoni. « High-speed Traffic Generation ». In : () (cf. p. 89).

[Rod15] Chris Rodriguez. « The expanding role of service providers in DDoS mitigation ». In : *Stratecast Perspectives and insight for Executives (SPIE)* 15.10 (2015), p. 1-10 (cf. p. 43).

[Rom12] Nuria Lloret Romero. « "Cloud computing" in library automation : benefits and drawbacks ». In : *The Bottom Line* 25.3 (2012), p. 110-114 (cf. p. 22).

[RWW12] Kui Ren, Cong Wang et Qian Wang. « Security challenges for the public cloud ». In : *IEEE Internet computing* 16.1 (2012), p. 69-73 (cf. p. 22, 23).

[Sac+10] Monika Sachdeva, Gurvinder Singh, Krishan Kumar et Kuldip Singh. « DDoS Incidents and their Impact : A Review. » In : *Int. Arab J. Inf. Technol.* 7.1 (2010), p. 14-20 (cf. p. 35).

[Sal+22] Loqman Salamatian, Scott Anderson, Joshua Matthews, Paul Barford, Walter Willinger et Mark Crovella. « Curvature-based analysis of network connectivity in private backbone infrastructures ». In : *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 6.1 (2022), p. 1-32 (cf. p. 32).

[San+15a] Jose Jair Santanna, Roland van Rijswijk-Deij, Rick Hofstede, Anna Sperotto, Mark Wierbosch, Lisandro Zambenedetti Granville et Aiko Pras. « Booters—An analysis of DDoS-as-a-service attacks ». In : *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE. 2015, p. 243-251 (cf. p. 35).

[San+15b] José Jair Santanna, Roland van Rijswijk-Deij, Rick Hofstede, Anna Sperotto, Mark Wierbosch, Lisandro Zambenedetti Granville et Aiko Pras. « Booters—An analysis of DDoS-as-a-service attacks ». In : *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE. 2015, p. 243-251 (cf. p. 63).

[Sau14] Molly Sauter. *The coming swarm : DDOS actions, hacktivism, and civil disobedience on the Internet*. Bloomsbury Academic, 2014 (cf. p. 38).

[SG10] Seungwon Shin et Guofei Gu. « Conficker and beyond : a large-scale empirical study ». In : *Proceedings of the 26th Annual Computer Security Applications Conference*. 2010, p. 151-160 (cf. p. 38).

[SH17] Aishwarya Soni et Muzammil Hasan. « Pricing schemes in cloud computing : a review ». In : *International Journal of Advanced Computer Research* 7.29 (2017), p. 60 (cf. p. 22, 23).

[Sha+15] Alireza Shameli-Sendi, Makan Pourzandi, Mohamed Fekih-Ahmed et Mohamed Cheriet. « Taxonomy of distributed denial of service mitigation approaches for cloud computing ». In : *Journal of Network and Computer Applications* 58 (2015), p. 165-179 (cf. p. 45).

[Sha+19] Iman Sharafaldin, Arash Habibi Lashkari, Saqib Hakak et Ali A Ghorbani. « Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy ». In : *2019 International Carnahan Conference on Security Technology (ICCST)*. IEEE. 2019, p. 1-8 (cf. p. 63).

[Sho+18] Tanishka Shorey, Deepthi Subbaiah, Ashwin Goyal, Anuraag Sakxena et Alekha Kumar Mishra. « Performance comparison and analysis of slowloris, goldeneye and xerxes ddos attack tools ». In : *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE. 2018, p. 318-322 (cf. p. 36, 37).

[SKS15] Raman Singh, Harish Kumar et RK Singla. « A reference dataset for network traffic activity based intrusion detection system ». In : *International Journal of Computers Communications & Control* 10.3 (2015), p. 390-402 (cf. p. 63).

[SLG18] Iman Sharafaldin, Arash Habibi Lashkari et Ali A Ghorbani. « Toward generating a new intrusion detection dataset and intrusion traffic characterization. » In : *ICISSp* 1 (2018), p. 108-116 (cf. p. 63).

[Som+16] Gaurav Somani, Manoj Singh Gaur, Dheeraj Sanghi et Mauro Conti. « DDoS attacks in cloud computing : Collateral damage to non-targets ». In : *Computer Networks* 109 (2016), p. 157-171 (cf. p. 34).

[Son+11] Jungsuk Song, Hiroki Takakura, Yasuo Okabe, Masashi Eto, Daisuke Inoue et Koji Nakao. « Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation ». In : *Proceedings of the first workshop on building analysis datasets and gathering experience returns for security*. 2011, p. 29-36 (cf. p. 63).

[SP20] Rabi Subudhi et Debajani Palai. « Impact of internet use during COVID lockdown ». In : *Horizon J. Hum. & Soc. Sci* 2 (2020), p. 59-66 (cf. p. 13).

[SSP12] Ramin Sadre, Anna Sperotto et Aiko Pras. « The effects of DDoS attacks on flow monitoring applications ». In : *2012 IEEE Network Operations and Management Symposium*. IEEE. 2012, p. 269-277 (cf. p. 34).

[Sto+09] Brett Stone-Gross, Marco Cova, Lorenzo Cavallaro, Bob Gilbert, Martin Szydlowski, Richard Kemmerer, Christopher Kruegel et Giovanni Vigna. « Your botnet is my botnet : analysis of a botnet takeover ». In : *Proceedings of the 16th ACM conference on Computer and communications security*. 2009, p. 635-647 (cf. p. 35).

[SU14] Krushang Sonar et Hardik Upadhyay. « A survey : DDOS attack on Internet of Things ». In : *International Journal of Engineering Research and Development* 10.11 (2014), p. 58-63 (cf. p. 36).

[Sun+07] Wei Sun, Kuo Zhang, Shyh-Kwei Chen, Xin Zhang et Haiqi Liang. « Software as a service : An integration perspective ». In : *Service-Oriented Computing–ICSOC 2007 : Fifth International Conference, Vienna, Austria, September 17-20, 2007. Proceedings 5*. Springer. 2007, p. 558-569 (cf. p. 21).

[Tav+09] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu et Ali A Ghorbani. « A detailed analysis of the KDD CUP 99 data set ». In : *2009 IEEE symposium on computational intelligence for security and defense applications*. Ieee. 2009, p. 1-6 (cf. p. 63).

[TM11] Alok Tripathi et Abhinav Mishra. « Cloud computing security considerations ». In : *2011 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. IEEE. 2011, p. 1-5 (cf. p. 32).

[TSB08] Ciza Thomas, Vishwas Sharma et N Balakrishnan. « Usefulness of DARPA dataset for intrusion detection system evaluation ». In : *Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security 2008*. T. 6973. SPIE. 2008, p. 164-171 (cf. p. 63).

[VHS11] ARi Vasudevan, E Harshini et S Selvakumar. « SSENet-2011 : a network intrusion detection system dataset and its comparison with KDD CUP 99 dataset ». In : *2011 second asian himalayas international conference on internet (AH-ICI)*. IEEE. 2011, p. 1-5 (cf. p. 63).

[Wan+15] Bin Wang, Zhengwei Qi, Ruhui Ma, Haibing Guan et Athanasios V Vasilakos. « A survey on data center networking for cloud computing ». In : *Computer Networks* 91 (2015), p. 528-547 (cf. p. 32).

[Wan10] Zhitao Wan. « Cloud Computing infrastructure for latency sensitive applications ». In : *2010 IEEE 12th International Conference on Communication Technology*. IEEE. 2010, p. 1399-1402 (cf. p. 23).

[Wei16] Joe Weinman. « Hybrid cloud economics ». In : *IEEE Cloud Computing* 3.1 (2016), p. 18-22 (cf. p. 25).

[Whe+14] Charles Wheelus, Taghi M Khoshgoftaar, Richard Zuech et Maryam M Najafabadi. « A Session Based Approach for Aggregating Network Traffic Data–The SANTA Dataset ». In : *2014 IEEE International Conference on Bioinformatics and Bioengineering*. IEEE. 2014, p. 369-378 (cf. p. 63).

[WLZ10] Li Wei-Min, Chen Lu-Ying et Lei Zhen-Ming. « Alleviating the impact of DNS DDoS attacks ». In : *2010 Second international conference on networks security, wireless communications and trusted computing*. T. 1. IEEE. 2010, p. 240-243 (cf. p. 38).

[WSZ08] Ping Wang, Sherri Sparks et Cliff C Zou. « An advanced hybrid peer-to-peer botnet ». In : *IEEE Transactions on Dependable and Secure Computing* 7.2 (2008), p. 113-127 (cf. p. 34).

[Xia+15] Peng Xiao, Wenyu Qu, Heng Qi et Zhiyang Li. « Detecting DDoS attacks against datacenter with correlation analysis ». In : *Computer Communications* 67 (2015), p. 66-74 (cf. p. 41).

[XKA16] KF Xylogiannopoulos, Panagiotis Karampelas et Reda Alhajj. « Real time early warning DDoS attack detection ». In : *Proceedings of the 11th International Conference on Cyber Warfare and Security*. Academic Conferences et Publishing International Limited Montreal, QC, Canada. 2016, p. 344-351 (cf. p. 43).

[XX16] Colin Ting Si Xue et Felicia Tiong Wee Xin. « Benefits and challenges of the adoption of cloud computing in business ». In : *International Journal on Cloud Computing : Services and Architecture* 6.6 (2016), p. 01-15 (cf. p. 23).

[Yad22] Shashank Yadav. « Political Propagation of Social Botnets : Policy Consequences ». In : *arXiv preprint arXiv :2205.04830* (2022) (cf. p. 34).

[Yeg+19] Bahador Yeganeh, Ramakrishnan Durairajan, Reza Rejaie et Walter Willinger. « How cloud traffic goes hiding : A study of Amazon's peering fabric ». In : *Proceedings of the Internet Measurement Conference*. 2019, p. 202-216 (cf. p. 32).

[Zhe+12] Zibin Zheng, Xinmiao Wu, Yilei Zhang, Michael R Lyu et Jianmin Wang. « QoS ranking prediction for cloud services ». In : *IEEE transactions on parallel and distributed systems* 24.6 (2012), p. 1213-1222 (cf. p. 32).

[Zho+17] Lu Zhou, Mingchao Liao, Cao Yuan, Haoyu Zhang et al. « Low-rate DDoS attack detection using expectation of packet size ». In : *Security and Communication Networks* 2017 (2017) (cf. p. 112).

[Zit03] Jonathan ZITTRAIN. « Internet points of control ». In : *The Emergent Global Information Policy Regime*. Springer, 2003, p. 203-227 (cf. p. 29).

[ZJ22] Jing ZOU et Cheng JIAN. « Does cloud computing improve team performance and employees' creativity ? » In : *Kybernetes* 51.2 (2022), p. 582-601 (cf. p. 13).

[ZLZ10] Yi ZHANG, Qiang LIU et Guofeng ZHAO. « A real-time DDoS attack detection and prevention system based on per-IP traffic behavioral analysis ». In : *2010 3rd international conference on computer science and information technology*. T. 2. IEEE. 2010, p. 163-167 (cf. p. 41).

[ZM09] Hossein Rouhani ZEIDANLOO et Azizah Abdul MANAF. « Botnet command and control mechanisms ». In : *2009 Second International Conference on Computer and Electrical Engineering*. T. 1. IEEE. 2009, p. 564-568 (cf. p. 34).

[ZZY17] Boyang ZHANG, Tao ZHANG et Zhijian YU. « DDoS detection and prevention based on artificial intelligence techniques ». In : *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*. IEEE. 2017, p. 1276-1280 (cf. p. 41).