République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Saad Dahleb Blida

Faculté des sciences

Département informatique

Mémoire de fin d'études

Pour l'obtention du diplôme de master en informatique

Spécialité : Sécurité de système d'information

Thème :

MISE EN ŒUVRE D'UN SYSTEME DE MONITORING POUR LA CYBER SECURITE

Mémoire présenté par :

- Hakimi Yacine

Promotrice : Mme Djeddar Afrah Encadré par : Dr. Zeghache Linda

Soutenu devant le jury composé de :

Mr. Douga **Président**

Mlle. Yekhlef **Examinateur**

Date de la Soutenance le : **juillet2019**

2018/2019

Remerciement

Je remercie en premier lieu, Allah tout puissant, de m'avoir accordé le courage et la volonté, pour achever ce travail.

J'adresse mes plus profonds remerciements à mes encadreurs Mme Zeghache Linda et Mme Djeddar Afrah qui ont assuré l'encadrement de ce travail au jour le jour et dont la disponibilité, la qualité des conseils et l'aide m'ont largement aidé à mener à bien cette étude.

Je tiens à exprimer toute ma reconnaissance envers tous mes enseignements du cycle primaire au cycle universitaire.

Il m'est très agréable d'exprimer toute ma sympathie à tous mes collègues, amis et ceux que j'ai côtoyés et appréciés.

Merci à mes parents, qu'ils voient ici le témoignage de ma profonde admiration et mon éternel amour.

Dédicaces

Je dédie ce modeste travail

A mes très chers parents

A mon très cher frère

A mes très chères sœurs

A mes amis de la promo

A tous ceux que j'aime et qui m'aiment

ملخص

مع زيادة التهديدات السيبرانية و محدودية وسائل المراقبة التقليدية ، أصبح استخدام أنظمة تلسكوب الشبكة أمرًا ضروريًا للحصول على معلومات موثوقة وفي الوقت المناسب لمواجهة التهديدات السيبرانية. تصف هذه الدراسة تصميم ونشر أول تلسكوب شبكة في الجزائر وتحليل البيانات التي تم جمعها بواسطة هذا التلسكوب.

يتم تخزين البيانات التي تم جمعها على مدى شهر واحد باستخدام مكدس Elasticsearch-Logstash-Kibana (ELK) الذي يسهل تحليل مجموعات البيانات الكبيرة.

يقدم هذا العمل عدة أنواع من تحليلات البيانات التي تم جمعها ، وهي: التحليل الأولي باستخدام NIDS ، وتنميط البيانات والتحليل الزمني باستخدام KIBANA والتحليل المتعمق باستخدام العنقدة. يتم استكشاف حركة المرور المجمعة بالتفصيل وإبرازها. كما تم تقديم مناقشة لحركة المرور المرصودة. و اقتراح نهج جديد لتصنيف حركة المرور واكتشاف الأنشطة غير العادية. لقد أظهرت نتائج العنقدة فعالية النهج المقترح ، وقدمت معلومات حول تهديدات معروفة وأخرى غير معروفة .

الكلمات المفتاحية: Cyber ، Darknet، تلسكوب الشبكة ، التهديدات ، الأمان ، الذكاء ، التهديدات السيبرانية ، NIDS العنقدة، ELK.

Abstract

With the increase of cyber threats and the limitations of traditional monitoring means the use of network telescope systems has become a necessity in order to obtain reliable and timely information to counter cyber threats. The present study describes the design and deployment of the first network telescope in Algeria and the analysis of data collected by this telescope.

Data collected over a period of one month is stored using the Elasticsearch-Logstash-Kibana (ELK) stack which facilitates the analysis of large datasets.

This work presents several types of analyzes of collected data, namely: preliminary analysis using NIDS, data profiling, temporal analysis using KIBANA and in-depth analysis using clustering. Collected traffic is explored in detail and highlighted. A discussion of observed traffic is also presented. A new approach to classifying traffic and detecting unusual activities is proposed. Clustering results have shown the effectiveness of the proposed approach, and provide information on known threats and other unknowns.

Keywords: Cyber, Darknet, Network Telescope, Threats, Security, Intelligence, Cyber-Threats, NIDS, Clustering, ELK

Résumé

Avec l'augmentation des cybermenaces et les limites des moyens de monitoring traditionnels l'utilisation des systèmes de télescope réseau est devenue une nécessité afin d'obtenir des informations fiables et au bon moment pour contrecarrer les cybermenaces. La présente étude décrit la conception et le déploiement du premier télescope réseau en Algérie et l'analyse de données collectées par ce télescope.

Les données collectées sur une période d'un mois sont stockées en utilisant le stack ELK(Elasticsearch-Logstash-Kibana) qui facilite l'analyse de grands ensembles de données.

Ce travail présente plusieurs types d'analyses des données collectées à savoir : une analyse préliminaire en utilisant les NIDS, le profilage de données, une analyse temporelle en utilisant KIBANA et une analyse approfondie en utilisant le clustering. Le trafic collecté est exploré en détail et mis en évidence. Une discussion relative au trafic observé est également présentée. Une nouvelle approche pour classifier le trafic et détecter les activités inhabituelles est proposé. Les résultats du clustering ont montré l'efficacité de l'approche proposée, et donnent des informations sur des menace connus et d'autres inconnues.

Mots clés : Cyber, Darknet, télescope réseau, Menaces, Sécurité, Intelligence, Cyber-menaces, NIDS,Clustering, ELK

Sommaire

Liste des figures

Liste des tableaux

Introduction générale

Le cyberespace est un nouvel univers sans frontières dans lequel tous les acteurs partagent de l'information et communiquent à travers Internet dans tous les domaines (des services bancaires à l'infrastructure gouvernementale), tout est contrôlé et exploité au moyen d'Internet.

La facilité, l'efficacité et la commodité de l'utilisation d'internet impliquent, dans une certaine mesure, d'exposer les utilisateurs à certaines menaces face aux pirates informatiques qui abusent de cette technologie devenue un outil peu coûteux pour générer des activités malveillantes telles que les virus, les vers, le déni de service (DoS), les scans et les botnets. Bien que leur concept ne soit pas nouveau, les techniques utilisées par ces menaces ont évolué ces dernières années ce qui rend la surveillance et l'investigation des cyberattaques plus complexe.

Le cyberespace oblige à repenser les normes de la sécurité. Lutter efficacement contre la cybercriminalité doit passer par une approche préventive qui consiste à rendre le cyberespace moins favorable à l'expression de la criminalité et à réduire les opportunités criminelles. Il faut élever le seuil de difficulté de réalisation des cyberattaques et accroître les risques pris par les criminels d'être identifiés, localisés et poursuivis.

Il faut pour cela collecter et analyser le flux du réseau, afin de détecter les attaques, d'identifier leur types, leur fréquences, leur sévérité et d'attribuer même ces attaques.

Le volume croissant du trafic sur les réseaux modernes rend la distinction entre le trafic légitime et le trafic malveillant plus complexe. Les services de surveillances traditionnels tels que le système de détection d'intrusion (IDS), le système de prévention d'intrusion (IPS), les pare-feux, les antivirus, les corrélateurs d'événements (SIEM) deviennent insuffisants [12].

Dans ce contexte, les systèmes de surveillances à base de leurres sont employés afin de piéger les pirates pour une détection précoce des menaces et la minimisation des dégâts.

Parmi les systèmes de surveillances à base de pièges on trouve le "Darknet" ou "télescope réseau". Son schéma consiste en une détection au moyen d'une analyse de trafic capturé par un serveur déployé dans des infrastructures de réseau avec des adresses IP publiques non attribuées[15].

L’objectif principal du présent projet, consiste à étudier le ***Darknet*** ou **telescope** en tant que système de surveillance à base de piège, et de mettre en place le premier telescope réseau en Algérie.

Le présent document est organisé comme suit :

Le premier chapitre présente la cybersécurité, les menaces, les logiciels malveillants, les politiques de sécurité ainsi les principaux mécanismes de sécurité.

Le second chapitre décrit les différents systèmes de surveillance basés sur les pièges du cyberespace (***Darknet, Honeypot, Greynet, Honeytokens et IP Gray Space***) et leurs taxonomies.

Le troisième chapitre est dédié à l'étude des Darknets comme systèmes de surveillance passifs à base de pièges, leurs techniques de déploiements, ainsi que de leurs avantages et inconvénients.

Le quatrième chapitre est consacré à la conception et la mise en œuvre d'un système de monitoring basé sur le darknet.

Enfin, le dernier chapitre présente les résultats obtenus en terme de qualité de données collectées par le darknet déployé et en terme de renseignement inférés par l'analyse de ces données.

Chapitre 1 : La Cybersécurité

1 Introduction :

Les réseaux actuels en perpétuelle évolution subissent un grand nombre d'attaques différentes, allant des intrusions système aux infections en passant par des outils d'attaque automatiques tels que les vers, les virus, les chevaux de Troie et les dénis de service (DoS). Yegneswaran et al [1] estiment qu'il existe 25 milliards de tentatives quotidiennes d'intrusion dans le monde et que cette activité continue d'augmenter.

Des outils tels que les systèmes de détection d'intrusion (IDS), les systèmes de prévention d'intrusion (IPS), les pare-feu, SIEM (Security information and event management) permettent de surveiller et d'analyser les activités de trafic réseau. Cependant, l'existence de techniques de cryptage et d'anonymisation rend le problème d'identification, de prévention et de surveillance des cyberattaques beaucoup plus difficiles.

Dans ce chapitre, nous examinerons certains des concepts de la cybersécurité, les principales menaces et cyberattaques, les solutions proposées pour protéger les utilisateurs, surveiller le trafic et détecter les attaques, les limites de ces solutions et la nécessité améliorer les solutions existantes.

2 Cyberattaque

Une cyberattaque est l'exploitation délibérée de systèmes informatiques, d'entreprises et de réseaux tributaires de la technologie. Les cyberattaques utilisent des codes malveillants pour modifier le code informatique, la logique ou les données, ce qui entraîne des conséquences perturbatrices qui peuvent compromettre les données et mener à des cybercrimes, comme le vol d'informations et d'identité. La cyberattaque est également connue sous le nom d'attaque de réseau informatique (CNA) computer network attack[2].

3 Anatomie d'une cyberattaque :

La plupart des attaques suivent le schéma illustré dans la figure 1 [3].

3.1 Cyber Scanning :

Ou la reconnaissance du réseau, est une étape indispensable dans toutes attaques organisées. C'est la première étape d'une tentative d'intrusion qui permet à un attaquant de récolter un

maximum de renseignements sur la cible, de localiser et d'exploiter à distance les systèmes vulnérables.

3.2 Enumération :

C'est le test des vulnérabilités découvertes pour identifier les points faibles qui permettent à l'attaquant d'avoir accès au système.

3.3 Tentative d'intrusion :

Le cybercriminel peut pénétrer dans le réseau ou utiliser des attaques avancées pour le rendre inutilisable.

3.4 Elévation du privilège :

Selon le modèle Microsoft STRIDE[1], l'élévation de privilège consiste, pour un utilisateur malveillant, à obtenir un niveau d'autorisation plus élevé que celui qui lui est normalement attribué.

3.5 Effectuer des tâches malveillantes :

Comme endommager ou voler des données.

3.6 Déployer des logiciels malveillants / porte dérobée :

Le cybercriminel installe des programmes malveillants sur le périphérique du point de terminaison cible pour créer ensuite une porte dérobée à travers laquelle plusieurs types de logiciels malveillants peuvent être téléchargés, permettant l'exécution de différentes attaques.

3.7 Supprimer les traces et les preuves et quitter :

C'est la dernière étape, les attaquants vont supprimer toutes les preuves de leur présence sur le réseau et les systèmes, ils utilisent souvent des virus et des vers pour détruire des preuves potentiellement incriminantes.

4 Les cyber menaces :

Plusieurs menaces existent sur Internet, nous présentons les principales parmi celles pouvant être détectées par les systèmes de surveillance du cyberespace.

[1]STRIDE est un modèle de menaces développé par Praerit Garg et Loren Kohnfelder chez Microsoft pour identifier les menaces de sécurité informatique.

4.1 Le scanning /probing :

C'est une activité de reconnaissance, elle est la première étape d'une cyberattaque. Son objectif est de découvrir les vulnérabilités sur une cible visée. Une fois qu'une machine est jugée vulnérable, l'attaquant tente de la contrôler ou de l'infecter en fonction de la vulnérabilité inférée. Les activités de scanning sont basées généralement sur les protocoles TCP, UDP ou ICMP.

Figure 1 .Anatomie d'une cyberattaque

Un réseau TCP / IP offre deux protocoles pour la couche de transport. L'un d'entre eux est le protocole UDP (User Datagram Protocol) sans connexion et non fiable. Une application utilise UDP fournit un transfert rapide et ignore les pertes de paquets. Le deuxième est le protocole de contrôle de transmission (TCP) fournit un service fiable et orienté connexion. Les en-têtes de paquet des protocoles TCP et UDP contiennent le numéro du port source et du port de destination.

Ces nombres sont des entiers non signés de 16 bits couvrant la plage de 0 à 65535. Les ports ouverts, également appelés ports d'écoute, sont des ports du système d'exploitation permettant d'établir la communication entre deux composants logiciels sur des ordinateurs différents. Il y a trois classes de ports :

- Ports réservés (numéros 0 à 1023) : Ces numéros sont réservés à des services et applications. Ils sont généralement utilisés pour les applications telles que HTTP (serveur Web)ou les processus du système d'exploitation, affectés aux services par l'autorité IANA (Internet Assigned Number Authority) [6].
- Ports inscrits : 1024-49151, principalement attribués par les processus des applications utilisateur, ces processus sont essentiellement des applications particulières qu'un utilisateur a choisi d'installer plutôt que des applications courantes qui recevraient un numéro de port réservé.
- Ports dynamiques ou privés : 49152-65535, utilisés par l'application utilisateur, ces ports sont généralement affectés de façon dynamique à des applications clientes lorsqu'une connexion à un service est initiée par un client.

4.2 Botnet (réseaux de zombies):

Les attaquants peuvent prendre le contrôle des ordinateurs connectés à Internet via des attaques directes ou indirectes, ces ordinateurs compromis sont appelés botnet. Les pirates informatiques distribuent et amplifient leurs attaques en utilisant les botnets. Un botnet est réquisitionné par un ou plusieurs botmasters pour réaliser des attaques telles que DDoS, spamming ou le scanning de port.

Le botmaster prend le contrôle du botnet par le biais d'un canal de commande et de contrôle (C&C), de cette façon les robots individuels deviennent partie intégrante du botnet et peuvent être utilisés pour effectuer des attaques coordonnées.

4.3 Exploit :

C'est un logiciel ou une séquence de commandes, utilisé afin d'exploiter une faille de sécurité d'un système d'information pour exécuter des actes malveillants.

4.4 Déni de Service (Denial of Service - DoS) :

Les attaques par déni de service sont des tentatives de rendre un ordinateur ou des ressources réseaux indisponibles, pour but d'empêcher les utilisateurs légitimes d'accéder à ces ressources. Il peut être lancé sous deux formes, le premier en envoyant un ou plusieurs paquets soigneusement conçus exploitant une vulnérabilité logicielle du système cible. Par exemple, l'attaque «Ping-of-Death»[2],la deuxième forme consiste à utiliser des volumes

[2]Envoie à un système cible un grand paquet de requêtes ping ICMP fragmenté en plusieurs datagrammes, ce qui peut provoquer le blocage ou le redémarrage de certains systèmes d'exploitation.

massifs de trafic inutile pour occuper toutes les ressources pouvant servir le trafic légitime, Lorsque le trafic d'une attaque DoS provient de sources multiples(en utilisant les botnet), il est appelé un déni de service distribué (DDoS) [4], DDoS attaque est en augmentation constante par exemple le 28 février 2018, le site d'hébergement de code de GitHub a été frappé par la plus grande attaque DDoS de l'histoire qui a culminé à 1,35 Tbps, et l'attaque DDoS la plus soutenue a duré 297 h au premier trimestre de cette année, selon le rapport du laboratoire Kaspersky [26].

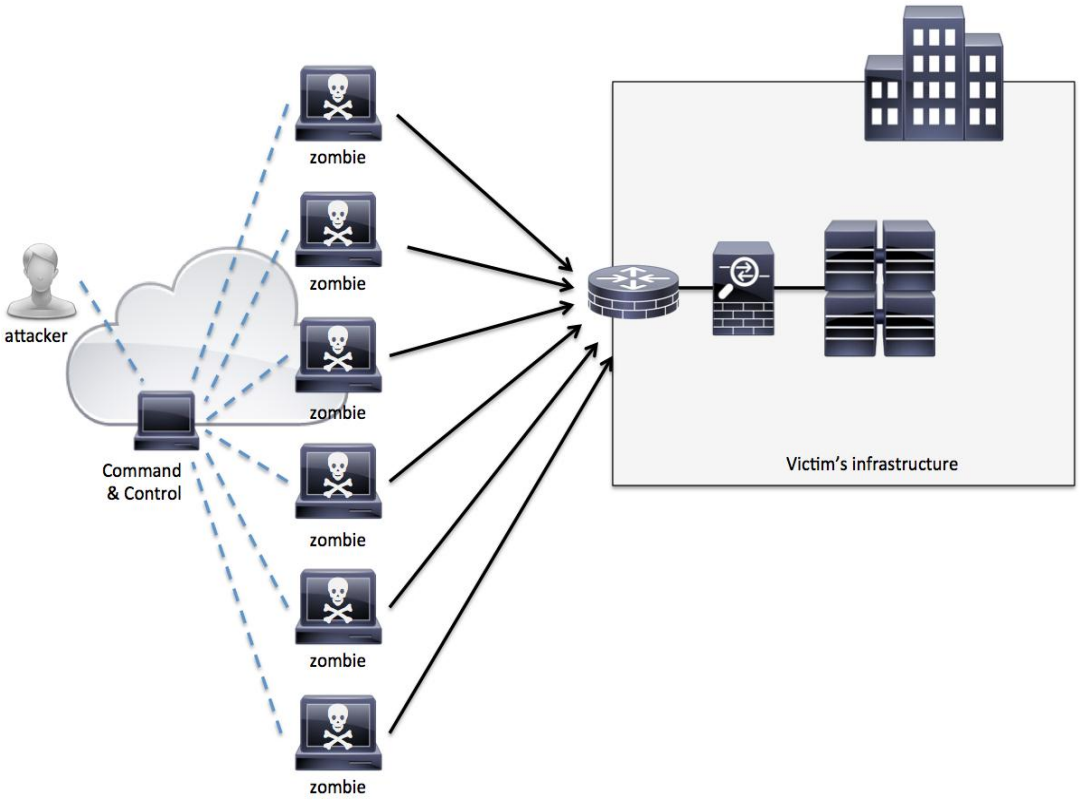


Figure 2. DDoS attaque

4.5 Distributed Reflection Denial of Service (DRDoS) :

DRDoS est un type spécial d'attaques DDoS. L'attaquant masque les sources du trafic d'attaque en utilisant des tiers (routeurs ou serveurs Web) pour relayer le trafic d'attaque à la victime, Ces tiers innocents sont aussi appelés réflecteurs, toute machine qui répond à un paquet entrant peut devenir un réflecteur potentiel.

Après que l'attaquant a pris le contrôle d'un certain nombre de "zombies", au lieu d'ordonner aux "zombies" d'envoyer directement le trafic d'attaque aux victimes, les "zombies" reçoivent l'ordre d'envoyer aux tiers du trafic spoofé avec comme adresse IP source l'adresse IP de la

victime. Les tiers enverront ensuite le trafic de réponse à la victime, ce qui constitue une attaque DRDoS [4].

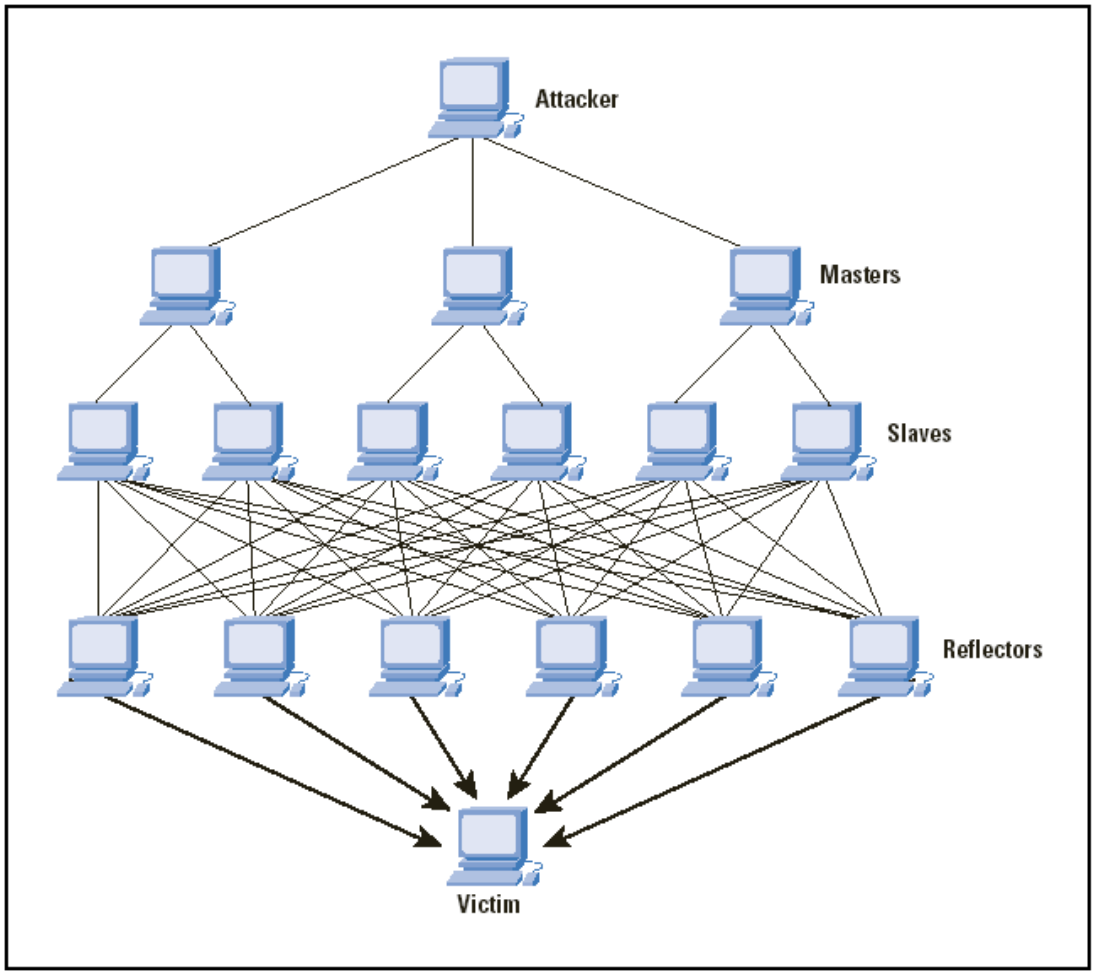


Figure 3.DRDoS attaque

4.6 Malware :

C'est un programme développé ou une partie d'un code conçu pour effectuer des activités malveillantes tels que les virus, les vers, les chevaux de Troie, etc. Certaines de ses caractéristiques peuvent inclure la propagation et la réplication.

4.7 Menaces persistantes avancées (Advanced Persistent Threats) :

Les APT font généralement référence à un groupe, tel qu'un gouvernement étranger, ayant à la fois la capacité et l'intention de cibler de manière persistante et efficace une entité. Ces cyberattaques possèdent des techniques de furtivité élevées et sont souvent spécifiques à une cible. Ils sont avancés car leurs opérateurs disposent de tout un éventail de techniques de collecte de renseignements. Les APT attribuent des priorités à des tâches spécifiques plutôt

que de rechercher de manière opportuniste des informations pour un gain financier ou autre. L'attaque est menée par une surveillance continue et une interaction afin d'atteindre les objectifs définis. Les attaques sont exécutées par des actions humaines coordonnées plutôt que par de simples morceaux de code automatisés. Leurs opérateurs sont généralement très compétents, motivés, organisés et bien financés [3].

4.8 Zero Day Attacks:

Ces attaques exploitent l'observation de vulnérabilités récemment découvertes mais non corrigées pour mener à bien leurs tâches malveillantes. Un certain nombre de mécanismes de détection ont été proposés pour se protéger contre ces attaques, mais ces attaques informatiques restent très dominantes et posent de graves problèmes [3]. Exemple : The Heartbleed Bug[3].

4.9 Forever-day vulnérabilités :

Ce sont les vulnérabilités qui prennent beaucoup de temps pour se fixer, ou ne sont jamais fixées. Certaines entreprises utilisent encore des programmes qui n'ont plus de mises à jour comme WindowsXP.

5 Cyberdéfense :

Pour atténuer les impacts de cybermenaces, de nombreuses entreprises et chercheurs proposent des mesures, réglementaires et techniques. Dans ce qui suit, nous nous concentrons sur les aspects techniques.

Les mesures techniques de cybersécurité comprennent les outils technologiques (logiciels et matériels) permettant de prévenir, détecter, atténuer et réagir aux cyberattaques.

5.1 Les normes de sécurité informatique :

La mise en œuvre de normes internationalement reconnues (Ex : ISO 27001).

5.2 Les mises à jour système :

Pour éviter les dénis de services applicatifs, on doit maintenir tous les logiciels de son système à jour puisque les mises à jour permettent souvent de corriger des failles logicielles, qui peuvent être utilisées par un attaquant.

[3]Heartbleed est une grave vulnérabilité de la célèbre bibliothèque de logiciels de chiffrement OpenSSL.

5.3 Les Antivirus :

Les antivirus sont des logiciels conçus pour identifier, neutraliser et éliminer des logiciels malveillants. L'antivirus analyse les fichiers stockés dans le disque dur et les fichiers entrants (fichiers téléchargés ou courriers électroniques) périodiquement, mais aussi la mémoire vive, l'analyse basée sur une base de données de signatures des logiciels malveillant et le comportement anormal de système.

5.4 Systèmes de détection du trafic malveillant :

Les techniques classiques pour la détection du trafic malveillant, de surveillance des réseaux et d'analyse du trafic réseau en général sont l'usage de pare-feu (fonctionne aussi comme un mécanisme de contrôle d'accès pour protéger la partie privée d'un réseau), les systèmes de détections d'intrusion IDS et IPS (Prévention /Protection contre les intrusions et non seulement la détection). Ces deux derniers aussi comme l'antivirus utilisent une base de signatures de vers et d'autres logiciels malfaisants.

5.5 Architecture DMZ (Demilitarized zone) [5] :

Une DMZ est un réseau situé entre le réseau local et Internet, il n'est ni à l'intérieur ni à l'extérieur du pare-feu. Il est accessible depuis les réseaux internes et externes.

Les règles de sécurité empêchent les périphériques externes de se connecter aux périphériques internes. Une zone démilitarisée est plus sécurisée que le réseau extérieur, mais moins sécurisée que le réseau intérieur. L'internet (réseau extérieur) est connecté à un pare-feu sur l'interface extérieure. Les utilisateurs et les serveurs qui n'ont pas besoin d'être accessibles à partir d'Internet sont connectés à l'interface interne. Les serveurs accessibles à partir d'Internet sont situés dans la zone démilitarisée.

Une DMZ a principalement deux objectifs :

- La première consiste à séparer les ressources d'accès public du reste du réseau.
- La seconde est de réduire la complexité.

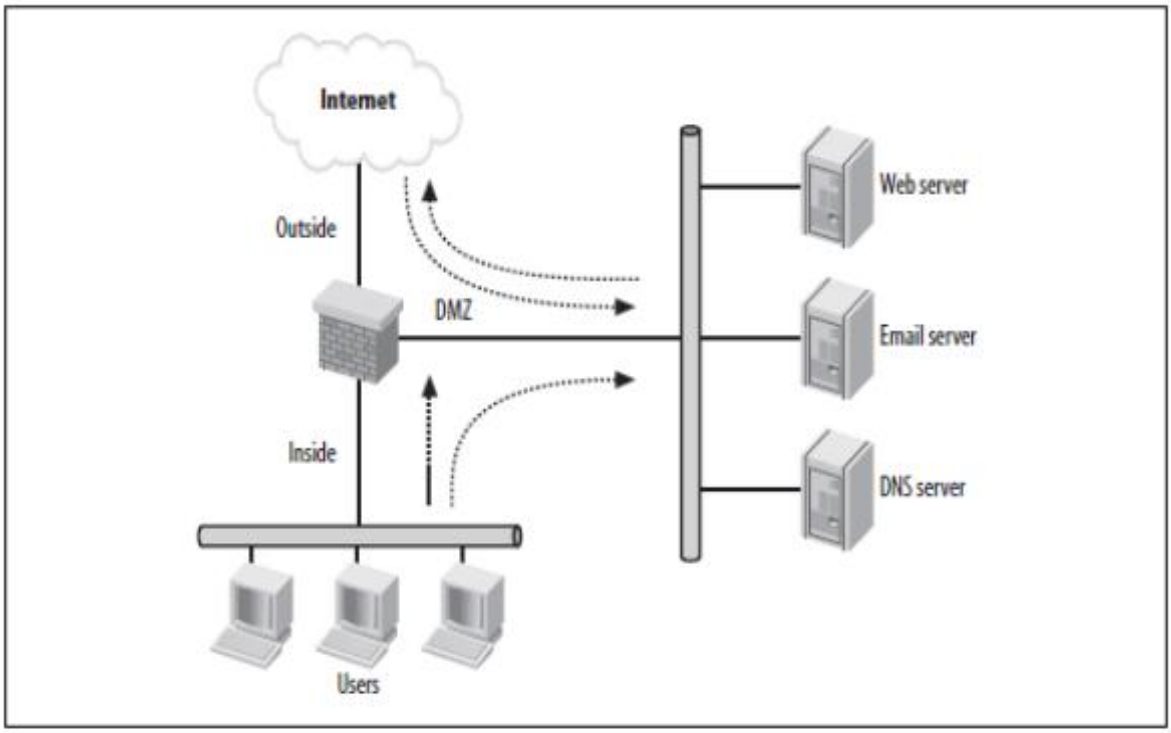


Figure 4.Architecture DMZ [5]

5.6 Cyber threat intelligence :

Cyber threat intelligence fait référence à la collecte de renseignements avant qu'un cyberattaquant ne cible un système victime. L'objectif est d'aider les organisations à comprendre et à atténuer les risques liés aux exploits Zero Day Attacks, aux menaces persistantes avancées (APT) et aux acteurs internes et externes de la menace. Cela permet aux organisations d'adopter une approche proactive en matière de cybersécurité et de prendre des contre-mesures préventives à l'avance, les données peuvent être obtenues de différentes sources [7].

5.7 Tactical Cyber threat intelligence :

Ces données proviennent de la surveillance en temps réel des systèmes (système de monitoring de cyberspace). Il s'agit d'événements en temps réel et d'informations relatives aux actions de l'adversaire à l'intérieur de l'organisation. Tactical threat intelligence (le renseignement tactique sur les menaces) est utilisé par les défenseurs pour s'assurer que leurs systèmes d'intervention en cas d'incident et leurs enquêtes sont préparés à cette tactique.

6 Conclusion :

Dans ce chapitre, nous avons présenté l'importance de la sécurité du cyber espace qui est devenu une partie importante de nos vies. Nous avons également fourni les mesures de protection pour la cybersécurité.

La collecte de renseignements ou le Cyberthreat intelligence, constituent également un aspect important de la protection, car il permet de prendre des mesures préventives les menaces inconnues (Zero Day Attacks, Forever-day vulnérabilités) et organisé (APT) .

Le succès de toute stratégie de cybersécurité est proportionnel à la quantité de renseignements sur les cyber-menaces disponibles pour l'analyse. Parmi les moyens efficaces de collecte d'informations sur les cybermenaces, on trouve les systèmes de monitoring du cyberspace qui feront l'objet du chapitre suivant.

Chapitre2 : Les systèmes de monitoring pour la cybersécurité

Les systèmes de monitoring pour la cyber sécurité permettent de rapporter l'état de fonctionnement du réseau afin de maintenir la disponibilité et détecter toute tentative d'intrusion qui risque de nuire au bon fonctionnement de l'entreprise. Ils donnent une visibilité sur l'ensemble du réseau par le filtrage, l'agrégation et l'analyse du trafic entrant et ce pour identifier le trafic malveillant. La détection de ce dernier au moment opportun, permet d'éviter les attaques ou du moins minimiser les dégâts. Néanmoins, cette tâche devient de plus en plus complexe avec l'augmentation du volume de données sur le réseau. A ce titre, plusieurs outils et techniques de surveillance sont développés et utilisés par les experts de sécurité.

Dans ce chapitre, nous allons passer en revue les systèmes de monitoring pour la cyber sécurité existants en accentuant sur les systèmes de monitoring à base de pièges.

1 Le trafic réseau malveillant

Le trafic réseau malveillant est tout type de trafic qui vise à atteindre un objectif menaçant l'un des trois piliers principaux de la sécurité informatique : la confidentialité, l'intégrité et la disponibilité.

La détection et l'identification du trafic malveillant est d'une importance cruciale. Ce dernier peut perturber le fonctionnement d'un réseau ou pire encore, il peut causer la perturbation de l'activité globale d'une organisation.

Le trafic réseau peut être surveillé et mesuré de différentes façons et méthodes. Cette surveillance comprend la classification du trafic en fonction de sa nature et de son importance.

Bien que la source du trafic ne soit pas toujours malveillante, le trafic peut devenir accidentellement malveillant lorsqu'elle viole involontairement l'un des trois piliers de la sécurité informatique.

Le trafic anormal est défini en fonction des politiques ou considérations d'une organisation. En termes de sécurité informatique, nous pouvons traiter de manière plus spécifique le trafic anormal en le reliant à des menaces bien définies, selon leur nature ou en utilisant des techniques ou des mécanismes spécialisés.

La plupart des méthodes de détection du trafic anormal sont basées sur la comparaison et l'analyse du comportement attendu de certains types de protocoles ou d'applications. Si le comportement et la forme du trafic sont bien identifiés en fonction de protocoles ou d'applications, alors tout forme ou comportement diffèrent représente un facteur permettant de l'identifier comme anormal.

En plus de ces techniques basées sur l'analyse de modèles, il existe d'autres techniqueset comme l'analyse des journaux.

2 Monitoring du cyberspace par les outils de détection de trafic malveillant

2.1 Analyseur de protocole (renifleurs)

Un analyseur de protocole est un matériel ou logiciel informatique utilisé pour fournir la capacité de capturer et d'enregistrer les paquets qui circulent sur le réseau afin de les analyser, sauvegarder et présenter sous une forme conviviale, il fonctionne de façon transparente pour le réseau.

Les analyseurs de protocole sont également mentionnés comme des analyseurs de réseaux et des analyseurs de paquets, ces outils utilisés pour l'analyse des problèmes de réseau, mais aussi pour la détection des tentatives d'intrusion réseau et le trafic malveillant, elles fournissent un bon ensemble de primitives pour filtrer et sélectionner le trafic en fonction de divers champs d'en-tête et valeurs de charge utile, elles sont considérées comme des outils d'analyse approfondie du trafic réseau en raison de leurs capacités de décoder les paquets et la majorité des principaux protocoles IP tels que TCP, UDP et ICMP, ainsi que pour certains protocoles de niveau supérieur tels que DNS(varie d'un outil à l'autre).

Elles peuvent être aussi utilisées à des fins malveillantes, de sorte qu'elles peuvent capturer le trafic et obtenir des informations privées.

L'utilisation de ces outils nécessite une connaissance approfondie des protocoles réseau et l'analyse de datagrammes (les techniques et la théorie nécessaires pour interpréter le contenu) pour une véritable exploitation, l'intérêt des analyseurs de paquets est de fournir les informations nécessaires sur le trafic qui peuvent être utilisées pour déterminer si le trafic est malveillant et de collecter des informations sur les menaces (Threat intelligence).

Beaucoup de systèmes IDS, IPS ou même les pares-feux ont ce type d'outils comme moteur pour capturer le trafic réseau à un niveau inférieur, ce qui leur permet de l'analyser en temps

réel, mais la différence réside dans le fait que généralement ne stockent que les paquets réseau par lesquels certaines alertes ont été détectées.

Les outils les plus connus sont basés sur la bibliothèque libpcap tel que: tcpdump[4], Windump[5], WireShark[6], TShark[7].

2.2 Pare-feu (Firewall) :

Un pare-feu est une combinaison de composants logiciels et matériels décidant si le trafic est autorisé à entrer ou à sortir d'un réseau local (LAN) ou d'un ordinateur [8].

La tâche principale d'un pare-feu est de bloquer certaines demandes de transfert de données. Le pare-feu prend ces décisions sur la base d'un ensemble de règles prédéfinies appliquées à chaque paquet, le filtrage des paquets peut avoir lieu à plusieurs couches. Cette surveillance de tout ce qui circule dans le réseau permet d'isoler, dans une certaine mesure, le trafic malveillant potentiel.

Il existe différents types de pare-feu à savoir :

- Par mode de fonctionnement :

  Le pare-feu utilise trois méthodes ou services de base pour protéger le réseau et peut être classé en fonction des services utilisés, les trois types (ou service) sont [9] :

  - Filtrage de paquets : examiner l'en-tête du paquet, vérifier l'adresse IP, le port ou les deux, et accorder et refuser l'accès sans apporter de modification (comprend pare-feu et pare-feu d'état).
  - Proxy de circuit (Circuit Proxy) : la principale différence entre le proxy de circuit et le pare-feu de filtrage de paquets réside dans le fait que le premier est le destinataire auquel tous les communicateurs doivent adresser leurs paquets, le circuit proxy remplace l'adresse d'origine (la sienne) par l'adresse de la destination souhaitée.
  - Application Proxy : plus complexe en opération que les deux précédents, il comprend le protocole d'application et les données, sur la base de ces

[4]https://www.tcpdump.org/
[5]WinDump est la version Windows de tcpdump
[6]https://www.wireshark.org/
[7]TShark est la version de ligne de commande de Wireshark

informations disponibles il prend des décisions, (Exemple : il peut authentifier les utilisateurs).

- ❖ Par l'installation :
  - Par feu hôte : Controller le trafic entrant et sortant d'un seul ordinateur.
  - Par feu réseau : Controller le trafic entrant et sortant d'un réseau, installé sur un routeur (logiciel) ou en tant que périphérique spécial (exemple CISCO ASA[8]).

Dans le cadre de la surveillance de la cybersécurité, le trafic malveillant détecté par le pare-feu peut être collecté et analysé à partir des fichiers journaux (logs) pour obtenir des informations sur les menaces potentielles (Threat intelligence).

2.3 Système de détection d'intrusion (IDS) :

Un IDS peut être une combinaison de logiciels et de matériels capable d'analyser le trafic réseau pour détecter et identifier le trafic malveillant. L'objectif de l'IDS est d'alerter mais pas d'agir (détection passive), les informations obtenues par l'IDS sur l'activité détectée est plus spécifique et détaillé que celle obtenue par un pare-feu.

La plupart des IDS essaient d'accomplir leur tâche en temps réel. Cependant, il existe aussi des IDS qui ne fonctionnent pas en temps réel, soit en raison de la nature de l'analyse qu'ils effectuent ou parce qu'ils sont destinés à Forensicanalysis[9][10].

Le fonctionnement de l'IDS est basé sur la détection de trafic malveillant au moyen de deux techniques :

- ❖ L'approche à base de signature :

  Elle s'appuie sur une base de signatures d'attaques qui consiste en un ensemble de définitions de modèles de menaces connus avec des caractéristiques spécifiques.

- ❖ L'approche comportementale :

  Aussi appelés détecteurs d'anomalies, présente l'avantage de pouvoir découvrir des attaques encore non répertoriées. Ils fonctionnent généralement en deux phases : une phase d'apprentissage qui définit certains critères de fonctionnement normal du système

[8]Adaptive Security Appliance est un périphérique de sécurité combinant des fonctionnalités de pare-feu, d'antivirus, de prévention des intrusions et de réseau privé virtuel (VPN).

[9]Forensicanalysis fait référence à une enquête détaillée visant à détecter et à documenter le déroulement, les raisons, les coupables et les conséquences d'un incident de sécurité.

ou du réseau, et une phase de détection, lorsqu'un certain trafic réseau détecté ne correspond pas aux critères définis, l'IDS peut l'identifier comme un trafic malveillant.

L'IDS présente deux principaux problèmes : les faux positifs et les faux négatifs.

- Faux positifs : ce sont les événements signalés sous forme d'alertes, mais qui ne sont pas en réalité du trafic malveillant.
- Faux négatifs : il s'agit des événements malveillants apparus sur le réseau mais ne sont pas signalés par l'IDS.

Les faux négatifs peuvent comporter un plus grand risque puisque, d'un certain point de vue, il est plus commode de détecter quelque chose qui n'existe pas que de ne pas détecter quelque chose qui existe réellement et qui est malveillant. Les faux positifs peuvent aussi être contreproductif surtout pour les activités de monitoring de cyberspace et l'analyse de trafic malveillant pour le threat intelligence.

2.3.1 Classification des IDS :

Les IDS peuvent être classés selon plusieurs critères tels que l'emplacement, les approches de surveillance et les techniques d'analyses [11] :

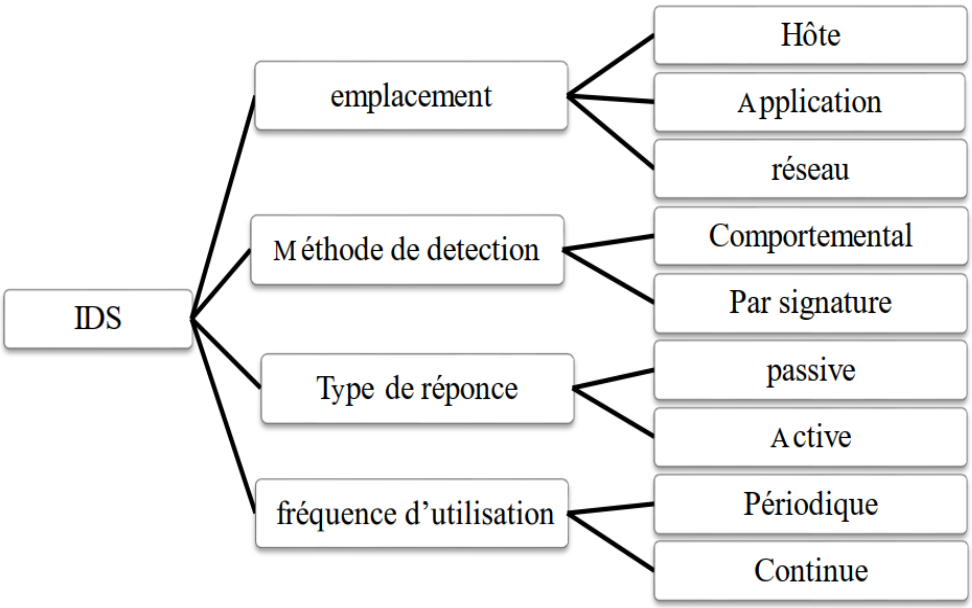


Figure 5. Caractéristiques des IDSs

La façon la plus commune pour classifier les IDS est de les regrouper par emplacement :

- IDS basé sur l'hôte (HIDS) : Les systèmes de détection d'intrusion basés sur l'hôte ou HIDS (Host IDS) analysent le trafic réseau qui entre et sort de cet équipement, les changements dans le système de fichiers ainsi que l'activité du système en général(figure 6).
- IDS réseau (NIDS) : Il analyse le trafic d'un réseau, il reçoit le trafic de tous les ordinateurs connectés au réseau (figure 7).
- IDS hybride : Les IDS hybrides sont basés sur une architecture distribuée en utilisant les deux types d'IDS, les informations sont envoyées à un seul serveur d'IDS (centralise l'information), Ce type de systèmes peut être utile dans les grands réseaux.

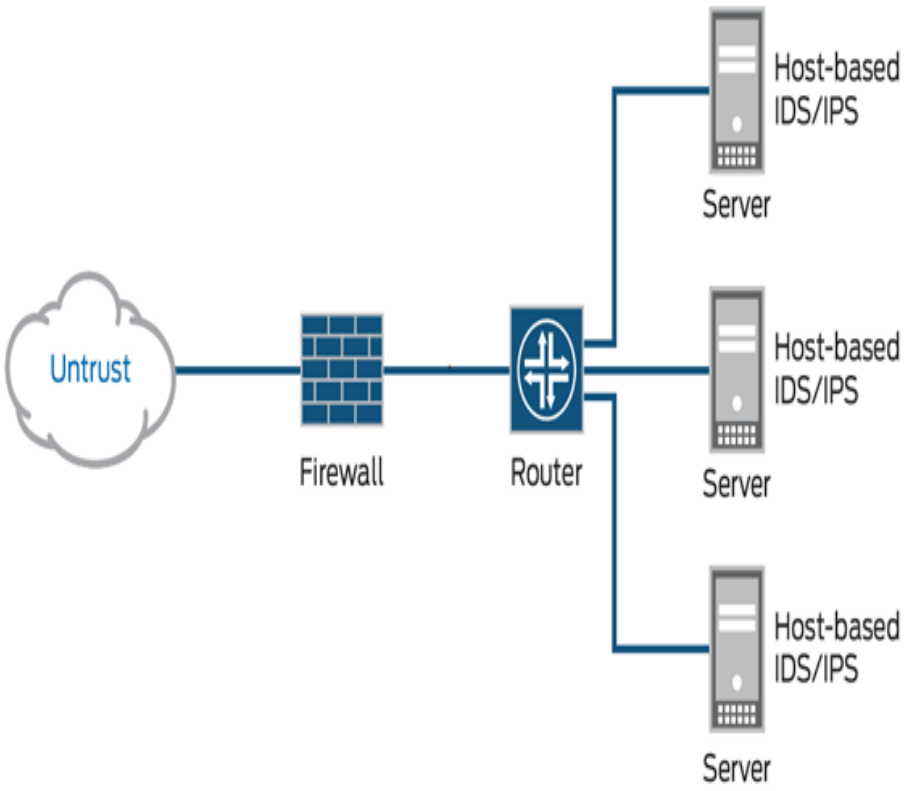


Figure 6.Schéma d'un IDS/IPS basé sur l'hôte

2.4 Système de prévention d'intrusion (IPS) :

Les systèmes de prévention des intrusions ou IPS, comme les IDS est une combinaison de logiciel et de matériel qui permet d'analyser le trafic réseau pour détecter et identifier le trafic malveillant. La principale différence avec l'IDS est que les systèmes de prévention des intrusions sont des dispositifs actifs qui ont la caractéristique d'agir de différentes manières en fonction des alertes détectées, et prévenir les attaques en temps réel ce qui implique que l'IPS dispose des capacités du pare-feu [12].

On pourrait penser que les IPS sont meilleurs que les IDS, face à une situation de plus en plus grave en matière de sécurité des réseaux, les systèmes classiques de pare-feu et de détection d'intrusion ne peuvent pas répondre aux besoins des utilisateurs [13], les entreprises internationales de sécurité et de recherche ont ouvert un grand débat sur la question de savoir si l'IDS est devenu obsolète, tandis que d'autres opinions ont exprimé le contraire, que la combinaison des technologies IPS, IDS et Firewall fournira une ligne de défense forte pour protéger les systèmes, avec ces trois technologies, on obtiendra une grande protection pour n'importe quel système[12].

En général Il existe trois types d'IPS :

- IPS réseau NIPS : Il analyse le trafic d'un réseau, il doit recevoir le trafic de tous les ordinateurs connectés au réseau. Il permet de surveiller le trafic à partir de plusieurs ordinateurs sur le réseau.
- IPS basé sur l'hôte HIPS : Il surveille le trafic et l'activité de l'ordinateur local.

2.5 Analyse des fichiers journaux (logs) :

L'analyse des logs est l'analyse des informations stockées dans les journaux d'un système afin d'identifier un événement ou une situation spécifique. Pour la détection du trafic malveillant ou de monitoring de cyberspace, une grande quantité de données peut être obtenue à partir des journaux de systèmes tels que les pare-feus, IDS/IPS, les équipements réseaux (commutateurs, routeurs, etc.), les journaux d'application qui utilise le réseau.

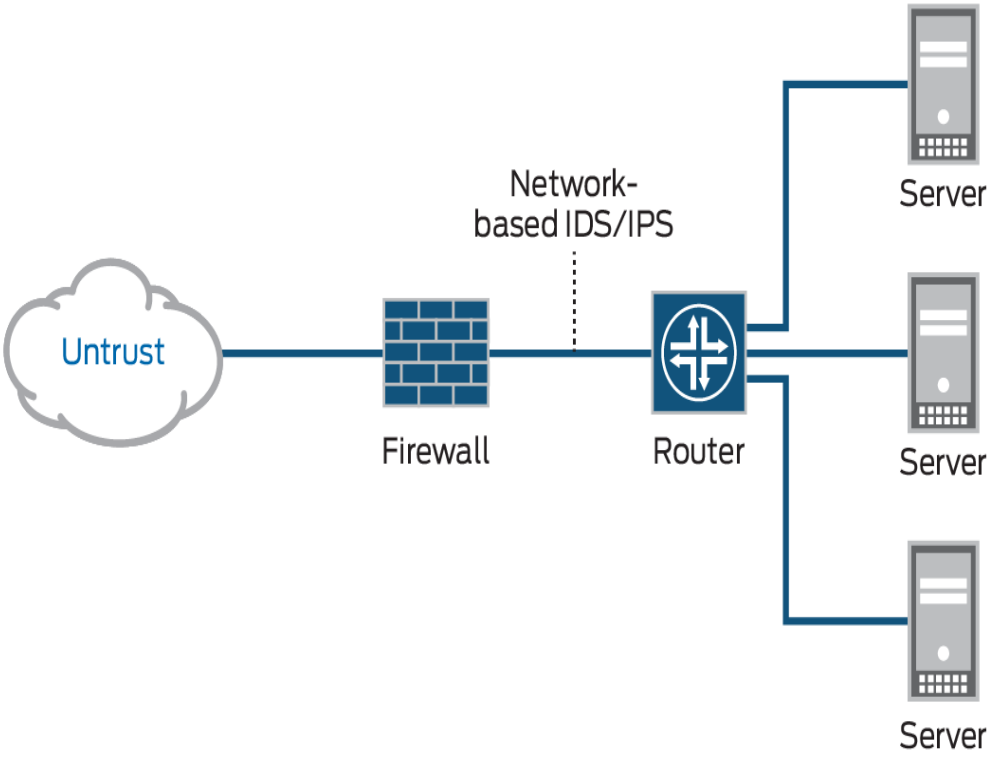


Figure 7. Schéma d'un IDS/IPS réseau

3 Les Systèmes de monitoring pour la cyber sécurité à base de piège

En raison de l'augmentation des cybermenaces et du nombre croissant d'attaquants qui inclut également des organisations criminelles et des pays (cyber-guerre); les méthodes usuelles de détection du trafic malveillant comme les pare-feu, l'IDS , et l'IPS conçus pour détecter les intrusions sont devenues eux-mêmes vulnérables aux nombreuses menaces conçues pour se défendre contre eux.

Compte tenu de ce qui précède et de l'énorme quantité de données pouvant être collectées par ces outils en fonction de la taille du réseau, et en plus des problèmes des faux positifs et négatifs, il est très difficile d'extraire des informations utiles et fiables sur les cyberattaques et les cybermenaces (Threat intelligence).

Afin d'obtenir des informations sur la cybermenace aussi fiable que possible, des technologies avec une approche d'analyse plus approfondie ont vu le jour ces dernières années à savoir, les systèmes de monitoring à base de pièges. Basées sur le concept d'émulation de services et de

réseaux, leur objectif principal est de recevoir, dans un environnement contrôlé, autant de trafic malveillant que possible pour être stocké et analysé.

Les systèmes de surveillance à base de piège sont des systèmes qui visent à piéger des adversaires en ligne. L'objectif est de collecter des informations sur les traces d'attaques et les activités telles que le scan de services vulnérables, la propagation de vers, les téléchargements de logiciels malveillants et d'autres activités de commande et contrôle telle que l'exécution de cyber-attaques DDoS via les Botnets [14]. La figure 8 décrit le concept de base de capteur de surveillance basé sur des pièges.

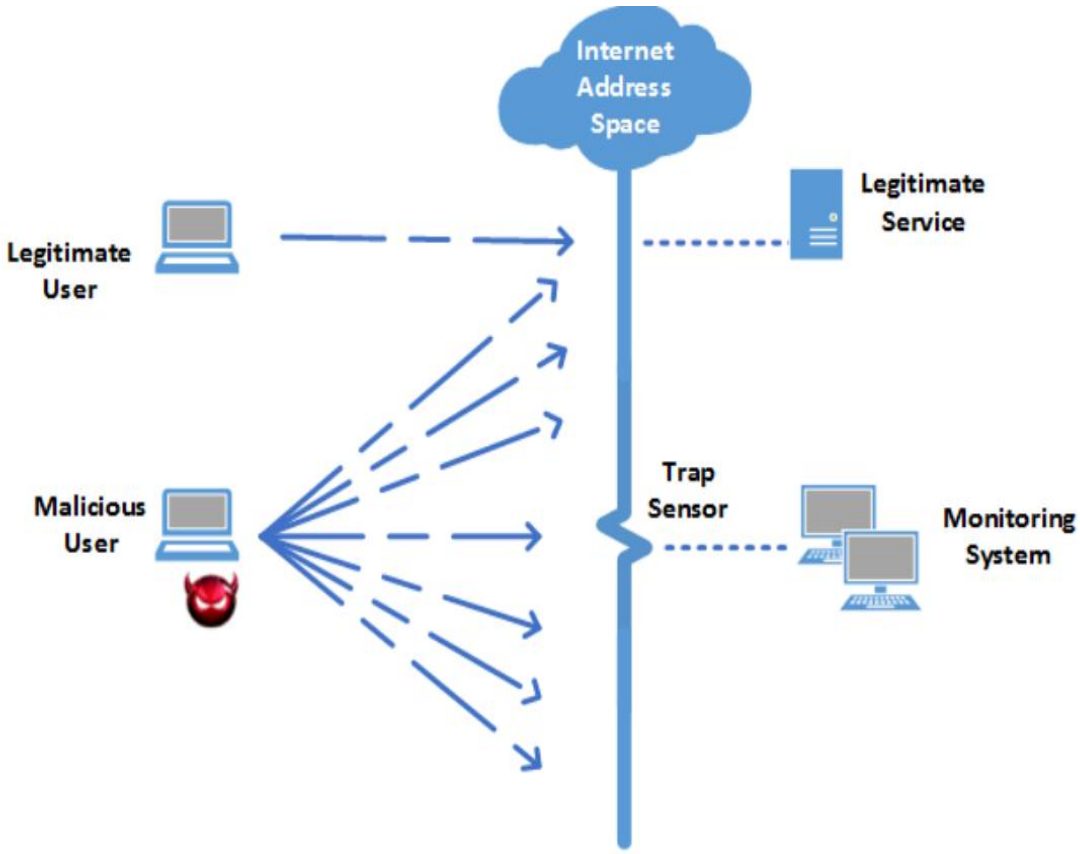


Figure 8.concept de base de capteur de surveillance basé sur des pièges [14]

Comme le montre la figure 8, un capteur de surveillance à base de pièges est déployé sur espace d'adressage Internet public pour attirer les utilisateurs malveillants. Dans certains cas, ces capteurs fonctionnent sur une adresse IP inutilisée, mais routable. Par conséquent, tout le trafic qui leur est destiné est considéré comme suspect et nécessite une investigation approfondie.

Le piège peut attirer les adversaires en exécutant des services vulnérables. Une fois les attaquants se sont connectés aux capteurs tout le trafic malveillant est transmis du capteur aux systèmes de surveillance pour une analyse plus approfondie.

Dans ce qui suit, nous allons présenter les principaux systèmes de surveillance à base de pièges soit***: les Darknets, les Honeypots, les Greynets, les Honeytokens et les IP Gray Space***

3.1 Darknet

Darknet, également connu sous le nom de télescope de réseau, Internet background Radiation ou blackhole, est un bloc d'espace d'adressage contigu fonctionnant sur des adresses IP inutilisées mais pouvant être routées. Les adresses IP non utilisées ne sont pas censées recevoir du trafic réseau, par conséquent, tout le trafic qui leur est destiné est susceptible d'être suspect [14]. Le Darknet est considéré comme une bonne source pour collecter des informations sur les cybermenaces (voir les détails dans le chapitre suivant).

3.1.1 Avantages et inconvénients :

Avantages :

Le principal avantage de cette approche est la sûreté car les capteurs Darknet sont passifs et n'interagissent pas avec les attaquants et donc ne sont pas menacés par l'attaque. La capture par Darknet est une surveillance passive (l'interaction avec la source d'attaque est nulle ou minimale).

Le déploiement du darknet est considéré comme facile par rapport aux autres systèmes de surveillance basés sur des pièges.

Inconvénients :

Le principal inconvénient du darknet est la collecte de données. En l'absence d'interaction avec l'adversaire, le darknet ne permet pas de détecter le scénario d'attaque complet, mais seulement la première étape de celui-ci.

3.2 IP Gray Space :

Ces adresses font référence à des périphériques qui ne sont attribués à aucun hôte pendant une période donnée (par exemple, 1 heure, 1 jour). En principe, l'espace IP Gray est très similaire à darknet. La seule différence est que les adresses d'espace gris IP sont inutilisées pendant une durée limitée, alors que les adresses darknet sont inutilisées de manière permanente [14].

IP Gray space peuvent être plus difficiles à détecter par un attaquant car ils sont actifs et fonctionnent en tant que machine normale à certaines périodes pour d'imiter des hôtes réguliers.

3.2.1 Avantages et inconvénients :

Il présente les mêmes avantages et inconvénients du darknet plus les suivants :

Avantages : plus de données seront collectées, telles que la communication entre programmes malveillants et réseaux de zombies.

Inconvénients : le fonctionnement en mode actif fournira aux adversaires des capacités d'attaque les serveurs de surveillance (capteurs), et le déploiement est plus difficile que le Darknet.

3.3 Honeypots (Les pots de miel) :

Les pots de miel sont définis de plusieurs manières, un pot de miel peut être défini comme un système informatique dont la valeur réside dans le fait d'être sondé, attaqué ou compromis, ou un système informatique ayant pour objectif d'attirer des pirates malveillants potentiels [16]. Les pots de miel nécessitent plus de ressources que le darknet, car ils interagissent pendant la communication.

Selon le niveau l'interaction, il existe 3 types principaux types de pots de miel à savoir, les pots de miel à faible interaction, moyenne interaction et haute interaction [14] :

Un pot de miel faiblement interactif dont la caractéristique principale est d'émuler les services d'un système réel afin de pouvoir interagir suffisamment avec des intrus. Étant une émulation, l'efficacité et la quantité d'informations recueillies dépendent de la complexité avec laquelle le pot de miel interagit. Cela implique un risque moindre mais les rend plus facilement détectables.

Un pot de miel interactif moyen est similaire au pot de miel interactif faible, mais avec des interactions supplémentaires et davantage de services émulés pour plus de capture et d'analyse de données.

Un pot de miel hautement interactif est un système informatique qui exécute un système d'exploitation complet vulnérable ou non corrigé et des applications telles qu'une version de système d'exploitation sur une machine virtuelle, ce qui implique un risque accru et donc la nécessité d'un système de contrôle externe pour le surveiller.

3.3.1 Honeynet :

Une collection de honeypots forme un honeynet.

3.3.2 Avantages et inconvénients :

Avantages : probablement c'est la meilleure source pour collecter des informations de sécurité car les traces peuvent suivre toutes les étapes de l'attaque, tels que le scan / probing, les exploits, les communications P2P et C & C, les activités de téléchargement de logiciels malveillants, le stockage de codes malveillants exécutables.

Inconvénients : L'interaction réelle du système avec l'attaquants les rend le plus vulnérables. Si les attaquants découvrent des services de surveillance suspects, ils peuvent bloquer la communication tracée ou même envoyer des informations non pertinentes pour décontenancer les enquêteurs.

3.4 Greynet :

C'est un réseau peuplé de darknet (inactif) et de pots de miel (actifs) dans le même espace d'adressage IP. En d'autres termes, greynet utilise à la fois le darknet (passif) et les honeypots (actif) sur le même espace de surveillance pendant la même période de temps.

L'objectif est de faire apparaître l'espace IP surveillé comme un piège plus attrayant pour l'attaquant.

Grâce à Greynet, il est possible de suivre les traces du darknet et du pot de miel. Déployer un greynet nécessite de déployer darknet et honeypot, le déploiement de Greynet est considéré comme le plus complexe de tous les systèmes de surveillance basés sur des pièges mentionnés ci-dessus.

3.4.1 Avantages et inconvénients :

Greynet présente les avantages et les inconvénients des systèmes de surveillance du darknet et du honeypot.

Avantages :

Un greynet est probablement l'espace le plus attrayant pour les adversaires du fait qu'il représente un réseau d'organisation typique ayant à la fois des hôtes actifs et inactifs.

La possibilité de suivre les effets de darknet et de honeypot fournit plus de données pour l'analyse, et donc plus d'informations sur les stratégies des cyberattaques.

Le greynet masque en quelque sorte les capteurs passifs, ce qui empêche les logiciels malveillants (tels que les virus à propagation automatique) d'éviter de toucher une adresse Greynet lors de la recherche de cibles d'infection [17].

Inconvénients :

Le greynet fournit une grande quantité et variété des traces disponibles pour l'analyse.

L'analyse est complexe du fait que le trafic provient de différents types de capteurs et donc mélange divers types d'attaques et de stratégies.

3.5 Honeytokens :

Les Honeytokens ont été introduits en 2003 par Augusto Paes de Barros, ce sont des pots de miel sans machines.

Les honeytokens sont des entités numériques, elles peuvent être n'importe quoi, du paramètre de l'application Web aux fichiers du système de fichiers, en passant par les données de la base de données, selon les besoins. Elle peut être implémentée sur toutes les couches possibles de l'application et les données d'application, ces données ne devraient jamais être utilisés dans un scénario d'utilisation normale, l'utilisation ou le changement de ces données signifie que quelqu'un a fait quelque chose qui n'est pas autorisé dans le contexte actuel de l'application [18]. En supposant que des pirates informatiques prennent le contrôle de ces informations et donc, par conséquent, les opérateurs de sécurité peuvent surveiller, suivre et retracer les activités de ces pirates.

Puisque le déploiement des Honeytokens ne nécessite pas d'entité physique, leur exécution est considérée comme simple et rapide.

3.5.1 Avantages et inconvénients :

Avantages :

Cette technologie est considérée comme facile à déployer et extrêmement efficace à faible coût.

La précision de Honeytokens dans la détection des utilisateurs malveillants est supérieure à celle d'un autre système de surveillance basé sur les pièges.

Le déploiement ne nécessite aucune de configuration et/ou d'installation de périphériques matériels.

Inconvénients :

Les honeytokens ne peuvent pas capturer toutes les informations réseau destinées à un espace d'adressage IP, car, comme mentionné précédemment, honeytokens est une entité numérique. Par conséquent, les Honeytoken ne parviennent pas à capturer les données concernant les activités de Scanning, la propagation du ver, etc.

3.6 Distribution d'espace d'adresse pour les systèmes de surveillance basé sur les pièges :

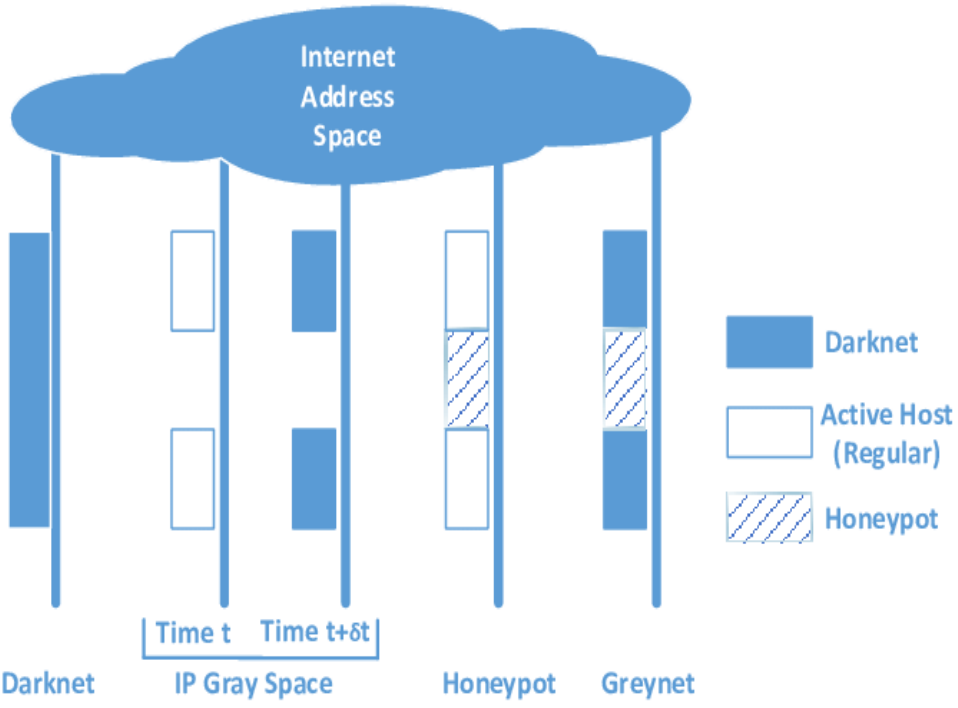


Figure 9. Distribution d'espace d'adresses [15]

L'espace d'adresse IP d'un darknet ne contient que les adresses inutilisées fonctionnant en mode passif (inactif).

IP gray space est similaire à darknet dans le temps (t+&t), toutefois, le même espace d'adressage était déjà actif dans une période précédente (t).

Honeypots peuvent fonctionner avec des hôtes actifs ou des hôtes passifs.

Greynet utilise les darknets et les honeypots dans le même espace d'adressage IP.

3.7 Comparaison :

Une ligne fine sépare les différentes formes des systèmes de surveillance de réseau basés sur des pièges. Les systèmes de surveillance sont comparés et classés en fonction de leur type, de leur niveau d'interactivité, de leur complexité, de la collecte de données et des aspects de sécurité [15]. Le tableau 1 présente le résumé de cette comparaison :

Tableau 1.Systèmes de surveillance basés sur des pièges –Comparaison [15]

Monitoring System	Type	Interactivity	Complexity	Data Collection	Security
Darknet	*passive IP*	*null*	*Low*	*low*	*secure*
IP Gray Space	*temporarily passive IP*	*null*	*Low*	*low*	*secure*
Low-Interactive Honeypot	*active IP*	*low*	*Low*	*low*	*vulnerable*
Medium-Interactive Honeypot	*active IP*	*medium*	*Medium*	*medium*	*vulnerable*
High-Interactive Honeypot	*active IP*	*high*	*High*	*high*	*vulnerable*
Greynet	*active/passive*	*low/medium/high*	*low/medium/high*	*low/medium/high*	*secure-vulnerable*

Interactivité : L'interactivité est la mesure du niveau d'interaction entre un adversaire et le système de surveillance.

La complexité : la complexité du déploiement la mise en place d'un greynet qui peut aussi fonctionner en pot de miel, peut-être plus complexe que le déploiement d'un simple darknet.

Collecte des données : La collecte de données mesure la quantité de données recueillies à partir du capteur du piège, un pot de miel hautement interactif qui stocke les communications bidirectionnelles peut recueillir plus d'informations qu'un darknet.

Sécurité : Mesure le niveau de sécurité du capteur mise en œuvre du côté de la surveillance. Par exemple, la sécurité du déploiement d'un pot de miel hautement interactif est élevée, car le capteur de surveillance pourrait être menacé d'être compromis.

3.8 Conclusion :

Comme nous l'avons mentionné dans ce chapitre, à mesure que les menaces à la cybersécurité augmentent, les chercheurs et les entreprises de sécurité de l'information ont besoin d'informations fiables sur ces menaces.

Le volume croissant du trafic sur les réseaux modernes rend la distinction entre le trafic légitime et le trafic malveillant plus complexe. Les services de surveillances traditionnels tels que les IDS et les par-feus s'avèrent insuffisants et donc le recours aux systèmes de surveillances à base de pièges devient une nécessité afin d'obtenir des informations fiables et au bon moment pour contrecarrer les cybermenaces.

Chaque type de systèmes de surveillance à base de pièges présente des avantages et des inconvénients. Dans notre étude nous allons explorer les systèmes de monitoring passifs à savoir les darknets et voir comment ils peuvent être déployés et exploités pour extraire les informations sur les menaces.

Chapitre 3 : Darknet : source pour la cyberintelligence

1 Introduction :

Outre l'augmentation du volume de données générées par les outils de surveillance de sécurité traditionnels et le nombre élevés des faux positifs, la communauté des chercheurs dans la cybersécurité rencontrent un problème majeur qui est l'indisponibilité des données réelles des entreprises telles que les logs des IDS, les logs des par-feus ou le trafic réel du réseau, en raison des règles de lois qui protègent la vie privée des personnes ou des entreprises. Par ailleurs, plusieurs études confirment que les menaces réseaux peuvent être observés à travers les systèmes de monitoring passifs à base de pièges qui sont les Darknets.

Ces systèmes de surveillance sont basés sur l'utilisation des adresses IP routables non utilisées pour attirer ou piéger les attaquants, cela permet de collecter des informations et extraire des renseignements sur les cybermenaces.

Dans ce chapitre, nous présentons le concept, le déploiement, l'exploitation et les schémas de mise en œuvre de la technologie Darknet. Ce chapitre fournit également un résumé de l'état de l'art sur les systèmes de surveillance darknet au cours des dernières années.

2 Définition :

Les Darknets, appelés aussi network telescopes, blackhole monitors, Sinkholes, ou background radiation monitors [19], sont un bloque contigu d'adresses IP publiques routables et non allouées. Ce terme est utilisé aussi pour décrire les systèmes de surveillance qui capturent le trafic réseau destiné à ces adresses en mode passif sans interaction et donc sans révéler aucune information sur eux-mêmes.

Comme ces adresses IP ne sont pas utilisées, elles représentent de nouveaux hôtes n'ayant jamais communiqué avec d'autres périphériques. Ces adresses inconnues ne sont pas censées recevoir du trafic. Par conséquent, tout trafic observé à destination de ces hôtes soulève des soupçons et nécessite donc une investigation.

Le grand avantage d'un Darknet est qu'en présentant de faibles pourcentages de faux positifs, les événements identifiés représentent une estimation générale de l'activité du trafic malveillant présente sur le réseau. Différents outils peuvent être utilisés pour obtenir ces informations en fonction du type d'analyse, du traitement et des résultats attendus.

3 Les types de menaces détectées par le Darknet :

Dans un environnement idéal, le trafic destiné aux adresses IP du darknet ne devrait pas exister, un tel trafic peut être des activités de scan envoyées par des virus/vers informatiques, ou du trafic de rétrodiffusion (backscatter) généré par des attaques DDoS ou DRDoS, Botnet C&C. Les adresses IP non utilisées peuvent recevoir aussi du trafic légitime du à un bogue logiciel ou à une mauvaise configuration.

3.1 Activités de scan (Probing/Scanning) :

Un ***darknet*** est en effet une approche efficace pour identifier diverses activités de scans ou d'analyse du réseau sur Internet [23]. La figure 10 fournit une illustration de la manière dont l'activité de scan est capturée par le système Darknet. La machine qui effectue le scan pourrait avoir été infectée auparavant par un ver[10] qui essaie de se propager ou qui participe peut-être à une analyse automatisée sur Internet. Certains de ces paquets peuvent atteindre le télescope et sont donc capturés par la suite [15].

[10]Un ver informatique est un logiciel malveillant qui se propage sur d'autres ordinateurs se répliquant et exploitant automatiquement les vulnérabilités en utilisant un réseau informatique, il n'a pas besoin de se lancer dans un autre programme ni d'être guidé par un utilisateur humain. Une fois installé, il peut, par exemple, télécharger un autre logiciel, tel qu'un outil d'accès à distance.

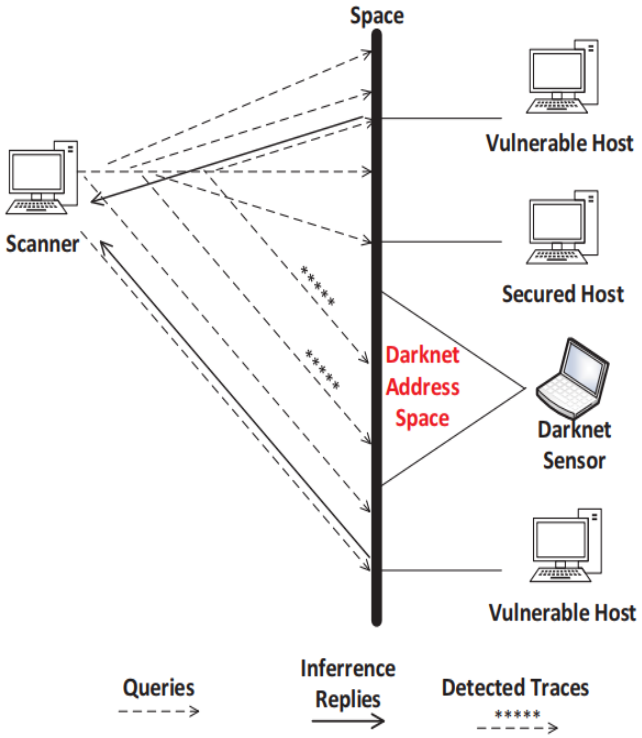


Figure 10. Activités de Probing [15]

3.2 Les attaques des deni de services distribués (DDoS)

Le but d'une attaque de déni e service est de rendre indisponible un réseau ou un serveur entier afin qu'il ne soit plus accessible par les utilisateurs légitimes. Une telle attaque génère du trafic de rétrodiffusion (backscatter), l'attaquant remplace son adresse par une adresse IP aléatoire avant d'envoyer les paquets à la victime. Donc les paquets de réponse sont acheminés vers la source usurpée et non pas à l'attaquant. Une telle adresse aléatoire pourrait appartenir au bloque d'adresses IP du darknet. La figure 11 montre comment le trafic de rétrodiffusion est capturé par le télescope.

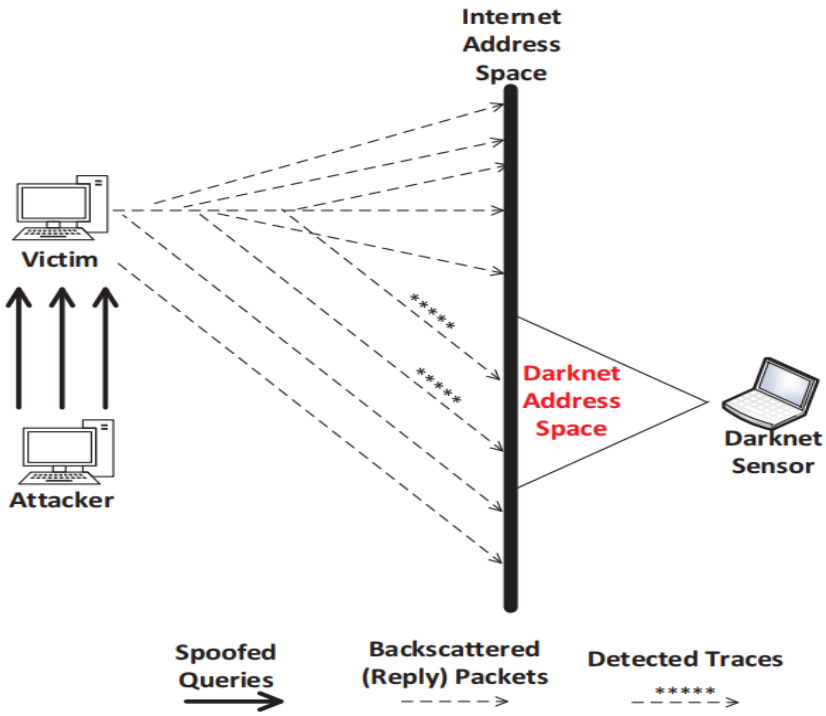


Figure 11Activités DDoS [15]

3.3 Les attaques DRDoS :

L'attaque DRDoS est un type spécial d'attaques DDoS. L'attaquant masque les sources du trafic d'attaque en utilisant des tiers (routeurs ou serveurs Web, ouvert résolvants DNS récursifs, serveurs UDP accessibles au public [24]) pour relayer le trafic d'attaque à la victime, les adversaires envoient des requêtes aux ces serveurs publics et usurper l'adresse IP d'une victime. Ces serveurs, à leur tour, inondent la victime de réponses valides et (sans le savoir) épuisent sa bande passante. Un darknet permet de déduire les attaques DRDoS[25]. La figure 12 illustre ce scénario. Généralement, l'attaquant pulvérise sur Internet des requêtes usurpées dans l'espoir d'atteindre autant d'amplificateurs ouverts que possible afin d'obtenir un facteur d'amplification élevé. Cela se produit dans le cas où les attaquants ne connaissent pas à l'avance les adresses IP des amplificateurs ouverts. Intuitivement, certaines de ces demandes atteindront le darknet et seront donc capturées [15].

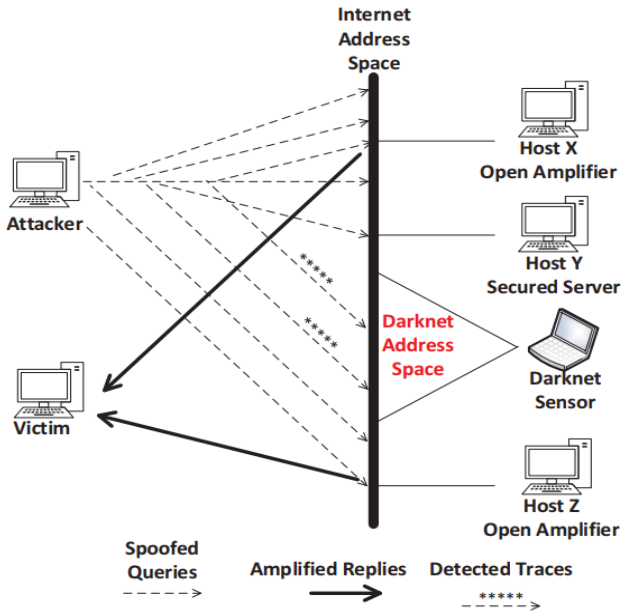


Figure 12 Activités DRDoS

4 Données Darknet :

En général les données darknet peuvent être classées dans l'une des trois grandes catégories suivantes [27] :

- BACKSCATTER : ou trafic rétrodiffusé, il résulte de l'utilisation de l'espace adresse surveillé pour l'usurpation d'identité. Le plus souvent sous forme d'analyses de decoy scans[11], d'attaques par déni de service (DoS) ou résultat d'hôtes mal configurés. Ce trafic se compose principalement de certaines classes de paquets ICMP et de paquets TCP contenant des indicateurs RST(réinitialisation), SYN (synchronisation) et ACK (accusé de réception).
- MISCONFIGURATION : Ce trafic pourrait être classé comme partiellement rétrodiffusé, ainsi que comme trafic potentiellement agressif, et résulte le plus souvent d'hôtes en ligne mal configurés.

- AGRESSIF / HOSTILE : La majeure partie du trafic observé sur le télescope du réseau darknet peut être classée comme agressive ou potentiellement hostile. Cela inclut , les activités DRDoS, les cas évidents d'analyses réseau se manifestent à la fois via l'analyse ICMP et TCP, et les paquets évidemment hostiles avec des charges utiles

[11]Cette technique utilise l'usurpation d'adresse, de sorte que plusieurs paquets similaires avec une adresse d'expéditeur différente sont également envoyés en même temps que les paquets de scan proprement dits., le destinataire n'aura aucun moyen de distinguer les paquets réels des paquets fictifs.

d'exploitation (ceux-ci ne sont visibles que dans le cas d'exploits basés sur UDP en raison de leur nature sans connexion). Le reste représente le trafic pouvant être groupé comme provenant de divers agents de numérisation automatisés tels que Internet Worms et les logiciels malveillants connexes.

Pour comprendre la nature des données ***darknet***, le tableau 2 fournit un aperçu de la distribution des protocoles. Il est montré que la majorité du trafic darknet est constitué de paquets TCP. Plusieurs faits peuvent expliquer la domination de TCP. Tout d'abord, TCP fournit diverses techniques d'analyse (par exemple, SYN, Fragmentation, SYN-ACK). Ensuite, Générer un scan TCP est généralement plus réalisable qu'avec UDP. Enfin, les cyberattaques bien connues ciblent spécifiquement les services TCP. Le tableau 3 énumère en outre les principaux protocoles d'application trouvés sur darknet.

Les tableaux 2 et 3 sont le résultat de l'analyse de l'ensemble de données est constitué de données ***darknet*** pures capturé au cours d'une période de cinq ans à partir d'un bloc d'adresses /8 [28].

Tableau 2. Distribution de protocoles [15]

Count	TCP	UDP	ICMP
Packet	*76.6%*	*19.9%*	*2.8%*
Bytes	*55.82%*	*40.82%*	*2.66%*

Tableau 3. Les principaux protocoles d'application trouvés

Port	Service
445	*microsoft-ds*
139	*NetBIOS*
4662	*eDonkey*
80	*HTTP*
135	*Endpoint Mapper*

5 Déploiement de Darknet :

Le déploiement d'un système de surveillance darknet nécessite une compréhension de la topologie du réseau local. Dans la mesure où un moniteur darknet observe le trafic vers les adresses non utilisées, le routeur en amont doit transférer les paquets non-distribuables au serveur darknet.

Cette section présente les techniques de déploiement du darknet et les recherches les plus importantes qui s'y rapportent.

5.1 Configuration :

Il existe trois techniques générales pour transférer des paquets vers un système de surveillance darknet [19] :

- **Envoyer au routeur des ARP Reply pour chaque adresse inutilisée :**

  Une méthode simple et utile lorsque on n'a pas l'accès au routeur, mais il faut renvoyer des arp reply périodiquement, car un compteur de cache ARP supprime les entrées ARP qui n'ont pas été utilisées pendant une certaine période de temps, cette période varie en fonction des périphériques et des systèmes d'exploitation.

- **Utiliser le routage statique pour un bloc d'adresses :** configurer le routeur pour acheminer statiquement un bloc d'adresses vers le système de surveillance darknet, cette technique est simple mais nécessite que le bloc d'adresses darknet soit spécifiquement mis de côté pour la surveillance.

- **Utiliser le routage statique pour transférer tous les paquets non utilisés d'un réseau d'organisation au système de surveillance darknet :** Une approche plus flexible consiste à acheminer tous les paquets destinés à des emplacements pour lesquels aucune adresse plus spécifique n'est configurée (et serait donc abandonnée) vers le système de surveillance. Par exemple : si une organisation a un bloc d'adresses /16, on peut créer une route statique vers le darknet pour l'ensemble du /16. Les paquets vers des adresses valides atteindront des préfixes plus spécifiques et seuls les paquets vers des blocs d'adresses non utilisées passeront par la route spécifiée pour /16.

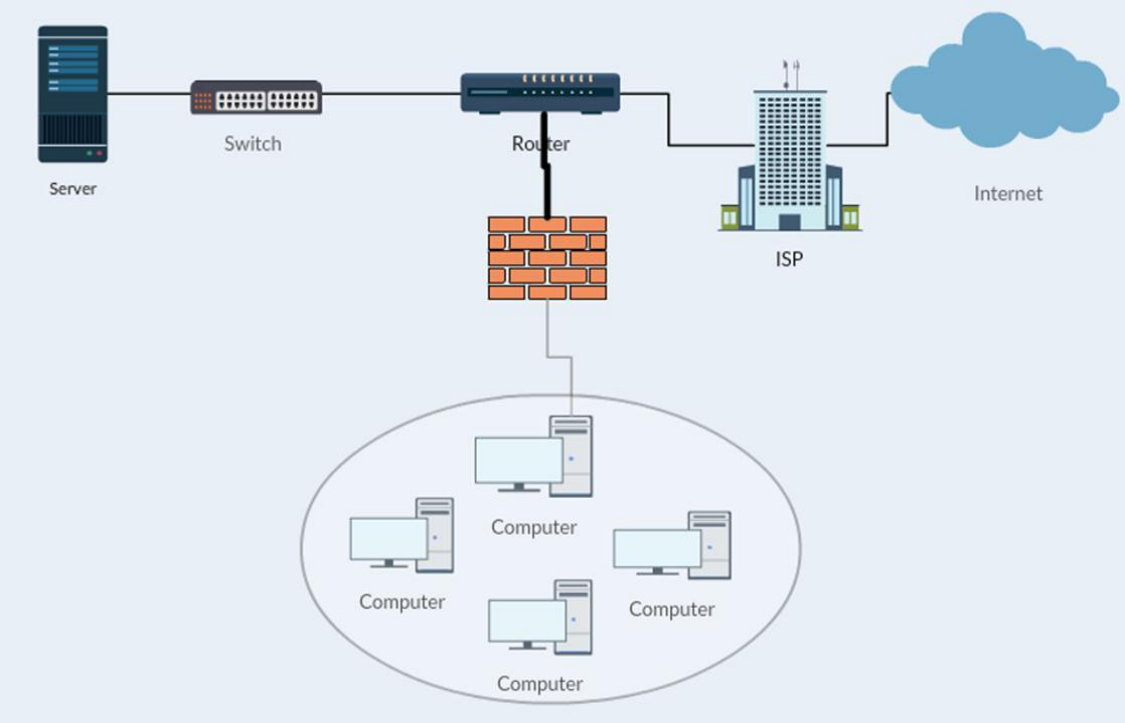


Figure 13. Déploiement d'un serveur Darknet

- Le sous-réseau entre le routeur et le système de surveillance doit utiliser des adresses privées.
- D'autres techniques doivent être utilisées comme le pare-feu, vlan, le routage interne selon l'architecture de réseau, pour garantir que le serveur de surveillance ne répond à aucun paquet.

5.2 Espace disque :

Comprendre les exigences de stockage d'un darknet est essentiel pour dimensionner correctement le système de surveillance, car le volume de trafic entrant peut être assez important. Ces exigences dépendent généralement du nombre d'adresses surveillées. Bailey et al (2006) [19] donne une estimation de ce trafic sur la base de l'analyse des données collectées par le Internet MotionSensor (IMS)[12]. En moyenne, un petit capteur /24 est susceptible de voir un taux de 9 paquets par seconde, un moniteur /16 verra environ 75 paquets par seconde, et un grand /8 moniteur plus de 5 000 paquets par seconde. Par conséquent nous pouvons estimer la taille de stockage dont nous avons besoin en multipliant le nombre de paquets × la taille du paquet × le temps de monitoring, nous pouvons plus tard recalculer nos besoins en fonction des résultats obtenus.

[12]Un réseau de capteurs Darknet distribués contrôlant 60 groupes d'adresses distincts dans 19 organisations réparties sur 3 continents.

En général pour le stockage à long terme, il est généralement préférable de conserver les données sur un système autre que le serveur darknet.

Tableau 4. Le nombre de paquets pour des blocs de réseaux darknet de tailles

Taille du capteur	Nbr de paquets enregistré par seconde
Petit capteur / 24	*09*
Taille moyenne / 16	75
Grand / 8	*5000*

5.3 Variantes Darknet :

Les variantes de ***darknet*** sont le déploiement des mécanismes de systèmes de surveillance basée sur des pièges utilisant des techniques semblables à ceux d'un ***darknet***, IP gray address space et greynet monitors (Nous les avons vus dans le chapitre 2), Harrop et al [17] utilisent le concept de greynet pour montrer comment un petit nombre d'adresses IP darknet mélangées avec des adresses IP actives peut augmenter l'efficacité de la détection des scans de réseau. Yu Jin et al. [20] utilisent le concept d'IP Gray Space en surveillance passive. Ils utilisent un algorithme heuristique pour identifier les adresses d'IP Gray Space. Polakis et al[21] proposent un système qui permet l'allocation dynamique d'un sous-réseau d'adresses IP non utilisé à l'utilisation d'un capteur de surveillance pour les réseaux qui utilisent DHCP.

5.4 La visibilité de Darknet :

L'activité malveillante observée par deux réseaux darknet de taille identique n'est presque jamais la même [17]. Ces différences ont tendance à dépendre de deux facteurs importants : l'emplacement d'un darknet et la manière dont un darknet répond aux paquets entrants.

La visibilité fournie par ***darknets*** dépend également fortement de la manière dont ***darknet*** répond aux paquets entrants. L'action la plus simple consiste à ne pas répondre du tout. Un ***darknet*** configuré passivement enregistre simplement tous les paquets qu'il observe et aucune autre action n'est entreprise. Cependant le darknet passif n'observe pas les données au niveau de l'application provenant d'hôtes qui tentent de se connecter via TCP, car toutes les transactions TCP valides nécessitent une négociation à trois voies qui doit être effectuée avant l'échange de données au niveau de l'application. Une technique de réponse active simple mais efficace consiste à répondre à un paquet TCP SYN avec des paquets TCP SYN-ACK [22].

6 L'analyse des données :

Différents outils et techniques de datamining, de statistiques et de visualisation ont été utilisées pour analyser les données darknet et nous pouvons résumer les recherches les plus importantes dans ce qui suit.

6.1 Profilage des données (Data Profiling) :

Englobe les travaux qui se concentrent sur la caractérisation des données darknet pour produire des statistiques et des aperçus, ces traveaux utilise des techniques telles que le filtrage de paquets, les séries chronologiques (time series).

Pang et al. [30] présentent une étude des caractéristiques générales du Darknet. Ils analysent à la fois les caractéristiques du trafic totalement non sollicité (analyse passive) et les détails du trafic engendrés par leurs réponses actives (analyse des activités). Irwin [31] discute des différences ainsi que des similitudes entre l'analyse des cinq capteurs darknet différents. Fukuda et al. [32] discutent des corrélations spatiales et temporelles entre le trafic indésirable par morceaux. Le but de leurs techniques est de déterminer s'ils peuvent estimer les propriétés statistiques du comportement global du trafic indésirable à partir de blocs d'adresses darknet plus petits. Ils ont constaté que la fluctuation du trafic darknet était presque aléatoire par rapport au trafic normal.

6.2 Filtrage et classification des données :

Plusieurs techniques et approches de classification et de filtrage des données darknet ont été proposées. Par exemple Glatz et al[33] ont proposé une approche basée sur un classificateur de trafic à sens unique pour faire la lumière sur la composition de darknet. Les auteurs ont constaté que ce trafic constitue la majorité de tout le trafic en termes de flux et peut être principalement attribué à des causes malveillantes. Wang et al [34] proposent une nouvelle approche pour filtrer le trafic darknet. Leur technique est basée sur la théorie de l'incertitude relative et est indépendante des configurations ou de la construction de bases de données. Les auteurs supposent que les données provenant des utilisateurs réguliers sont relativement certaines et non aléatoires. De plus, Cowie et Irwin [35] discutent des difficultés à générer du trafic d'entraînement pour l'analyse en intelligence artificielle (IA). Les auteurs mentionnent le problème de l'étiquetage précis des incidents connus de darknet.

6.3 Extraction de CyberThreatIntelligence à partir de données darknet :

Les chercheurs ont proposé de nombreuses techniques pour extraire divers types d'informations pertinentes à partir de données darknet. Divers chercheurs utilisent des séries

chronologiques et des méthodes statistiques pour établir le profil des menaces darknet. Harder et al [36] étudient les propriétés statistiques des adresses darknet de classe C depuis plus de trois mois. Les auteurs ont constaté que la majorité du trafic est basée sur peu de sources IP et d'adresses de destination. De plus des modèles tels que le scanning de port et la propagation de programmes malveillants [37, 38] peuvent être extraits. L'analyse de rétrodiffusion(backscatter) peut donner un aperçu des attaques sur Internet, en utilisant plusieurs techniques telles que les modèles mathématiques, le routage réseau, le filtrage des paquets et la visualisation. Par exemple [39,40] se concentrent sur la détection des menaces DDoS à l'aide greedyalgorithms. Darknet peut également être une source d'information sur l'activité des logiciels malveillants sur Internet. Moore et al [41] analysent le ver code-rouge, les auteurs ont géographiquement localisé et mesuré la population du ver et ont vérifié les (FAI) et les domaines de premier niveau affectés. Le ver Slammer a été étudié grâce aux télescopes réseau Darknet. Moore et al[42, 43] ont montré comment ce ver sélectionne ses victimes et ont expliqué les raisons de sa propagation rapide. Pour étudier les bontnetDagon et al [44] ont utilisé darknet pour comparer les taux de propagation des différents botnets, à prioriser la réponse et à prédire les infections futures des botnets. Les auteurs ont constaté que les fuseaux horaires jouent un rôle important dans la dynamique de croissance des botnets. Claude et al [45,46] ont proposé une nouvelle approche pour déduire et caractériser les activités DRDoS basées sur le DNS à grande échelle à travers l'espace darknet.

6.4 Mauvaise configuration des données (Data Misconfiguration) :

François et al. [47] à l'aide de l'analyse darknet ont montré que les réseaux déployés souffrent d'erreurs bien connues et d'une configuration erronée. Labovitz et al. [48] présentent une vaste étude sur les différences unilatérales d'accessibilité des fournisseurs de services Internet. Les auteurs se concentrent sur le darknet et la gamme de topologies accessibles à certains fournisseurs mais inaccessibles via un ou plusieurs réseaux concurrents, les résultats ont montré qu'Internet était effectivement partitionné et que darknet existait en grande quantité (5% des adresses Internet).

7 Les projets Darknet :

Au niveau mondial, il existe plusieurs projets de tailles et de caractéristiques différentes, certains plus complexes que d'autres, mais qui poursuivent un concept commun : la surveillance de l'activité du trafic réseau à l'aide de Darknet. Voici un aperçu général de certains de ces projets.

7.1 Projets darknet à grande échelle:

Certains d'entre eux, de par leur taille et leur capacité, sont devenus des références importantes pour consulter l'activité du trafic réseau et les tendances sur Internet et pour extraire divers types d'informations sur les cybers menaces.

7.1.1 Le télescope de réseau :

Le projet de télescopes de réseau UCSD[13] [49, 50] est un système proposé par des chercheurs du Centre d'analyse appliquée des données Internet (CAIDA)[14], qui consiste en un Darknet ayant le potentiel d'un réseau /8, soit environ 16millions d'adresses IP (l'une des plus grandes infrastructures mondiales de mesure et d'analyse du trafic Internet). Ce télescope a pour but de détecter les attaques par déni de service, la propagation des vers et la détection générale du trafic malveillant généré par les agents automatisés.

7.1.2 Système d'analyse active du niveau de menace (ATLAS[15]) :

Sous la direction d'Arbor Networks[16] [51], ce système de surveillance de réseau analyse collectivement les données traversant un réseau Darknet disparate afin de visualiser les activités malveillantes sur Internet et fournir aux clients d'Arbor services des informations relatives aux activités malveillantes telles que les exploits, le phishing, les logiciels malveillants et le botnet.

7.1.3 Le projet Darknet TEAM CYMRU :

L'équipe Cymru est une organisation spécialisée dans la recherche en sécurité Internet. L'un de ses projets est "The Darknet Project" [52] qui, comme ses pairs, est capable d'identifier les activités malveillantes sur Internet et de générer à son tour des statistiques de trafic pour savoir ce qui se passe sur le réseau. Il utilise des technologies telles que l'analyse des flux et l'analyse du trafic réseau.

L'infrastructure de ce projet se compose de 8 Darknets déployés dans différentes zones géographiques avec un total de 626 944 adresses IP.

7.1.4 Capteur de mouvement internet (IMS) :

Il s'agit d'un projet développé entre la société de sécurité Arbor Networks et l'Université du Michigan. Il s'agit d'un système mondial de surveillance des menaces Internet qui vise à

[13]University of California San Diego
[14] Center for Applied Internet Data Analysis
[15]Active ThreatLevelAnalysis System
[16]Arbor Networks est une société de logiciels fondée en 2000 et baséeà Burlington,Massachusetts, États-Unis, qui vend des logiciels de sécurité réseau et de surveillance de réseau

mesurer, classer et suivre les menaces Internet. Il dispose d'une infrastructure distribuée de capteurs situés à différents endroits couvrant des segments de réseau de /25 à /8 dans les réseaux académiques, commerciaux et FAI[22]. Son fonctionnement est basé sur différentes technologies de détection, y compris les honeypots, l'analyse de flux, l'analyse de charge utile, avec la possibilité de le faire en temps réel.

7.1.5 project de Network Incident analysis Center for Tactical Emergency Response (NICTER)[53]:

Est un système d'analyse d'incidents de réseau à grande échelle du centre japonais qui surveille principalement darknet. Il représente un système capable de capturer et d'analyser les programmes malveillants exécutables. L'identification de la propagation de programmes malveillants est l'objectif principal de ce projet.

7.2 Les projets à petite échelle :

Par exemple, Antonatos et al. [54] proposent HoneyHome, une partie du projet NoAH [55], une plate-forme de surveillance des adresses IP et des ports inutilisés pour l'extraction d'événements de sécurité à grande échelle. Ce système peu coûteux repose sur l'installation de capteurs sur des utilisateurs réguliers pour surveiller ces adresses IP et ports non utilisés. De plus, Daedalus [56], basé sur le projet NICTER, est conçu pour capturer les cyberattaques en temps quasi réel. Il existe aussi autres projets utilisant la surveillance passive, tels que le système d'acquisition de données par numérisation Internet (ISDAS) géré par le Centre de coordination CERT du Japon [57].

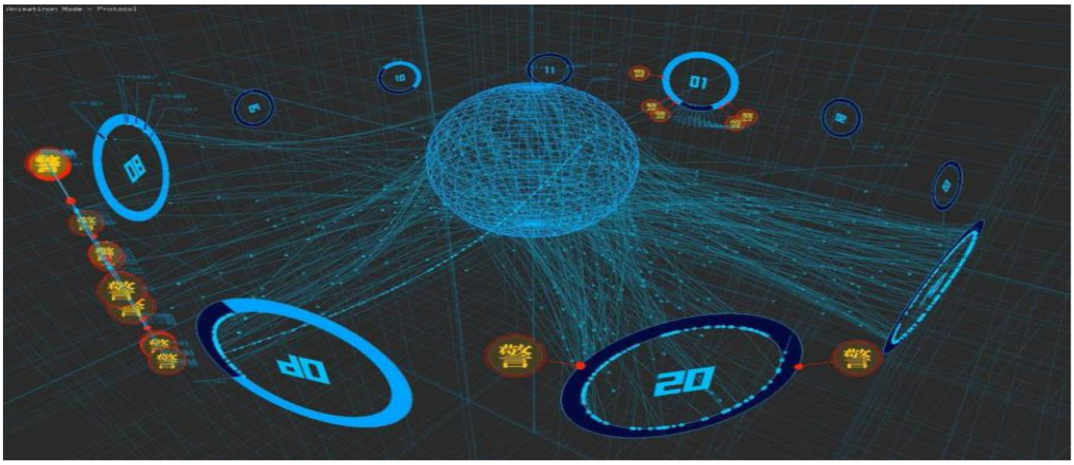


Figure 14. AperçuDAEDALUS-VIZ

7.3 Les projets en Afrique :

À notre connaissance, le premier et seul projet darknet en Afrique a été lancé en 2011 en Afrique du Sud : Rhodes University Network Telescope [58].

8 Visualisation Darknet :

Il existe beaucoup de recherche sur l'utilisation du trafic darknet dans la détection d'activités malveillantes en exploitant des techniques et des outils de visualisation.

Le et al. [59] proposent une nouvelle approche pour déduire le trafic réseau malveillant basée sur des concepts de la théorie des graphes tels que la distribution des degrés, les mesures de degré maximum et de distance. Les auteurs modélisent le trafic réseau à l'aide de la technique des graphiques de dispersion du trafic (TDG). De même, Joslyn et al. [60] proposent une nouvelle technique pour faciliter et visualiser des données à grande échelle. L'approche basée sur les graphes exploite les bases de données de routage réseau. Dans une autre contribution de visualisation, Krasser et al. [61] construisent un système de visualisation du trafic réseau capable d'analyser des données en temps réel en appliquant des techniques efficaces de visualisation des informations afin de réduire le taux de faux positifs et de faux négatifs. Fukuda et al. [62] proposent une technique permettant de détecter les activités de balayage dans le trafic darknet basée sur le traitement d'image appliquée à une image en deux dimensions qui représente un trafic indésirable. Harrop et Armitage [63, 64] décrivent un système dans lequel une technologie de moteur de jeu en 3D est utilisée pour permettre un contrôle de réseau collaboratif. L'approche proposée utilise des techniques d'interaction simplistes en traduisant les événements du réseau en activités visuelles.

9 Conclusion

Dans ce chapitre nous avons vu de près le déploiement des systèmes Darknet et leur fonctionnement, ainsi que les recherches les plus importantes concernant l'analyse et la visualisation des données Darknet et leurs exploitations pour la cyber intelligence. Nous avons présenté également les projets les plus importants dans ce domaine.

Compte tenu de l'absence de projets similaires en Algérie et de l'existence d'un seul projet en Afrique du Sud, l'espace d'adressage Internet en Afrique n'est pas surveillé et peut être considéré comme le plus susceptible d'être utilisé dans des activités malveillantes. Par conséquent, il est nécessaire d'aller vers des projets pour surveiller cet espace d'adressage.

Chapitre 4 : Conception et mise en œuvre d'un système de monitoring basé sur le Darknet

1 Introduction :

Ce quatrième chapitre présente la conception et la mise en œuvre d'un système de monitoring pour la cybersécurité pour CERIST. Il comprend toutes les considérations nécessaires à la mise en œuvre d'un Darknet dans le réseau CERIST, du matériel et des logiciels utilisés, aux bases de données. Notre vision de cette première version était d'établir un système qui serait simple à mettre en œuvre, facile à déployer, Scalable à l'avenir, tout en maintenant la sécurité du réseau utilisé. Le personnel informatique du centre s'inquiétait principalement de la possibilité d'attirer une attaque par déni de service dirigée contre le système de télescope et l'espace d'adresse associé, ce qui risquerait de compromettre la sécurité du centre.

Bien qu'il existe deux grands types de Darknet comme mentionné précédemment passive et active, pour mettre en œuvre le système de collecte du télescope de manière à minimiser tout risque potentiel, donc pour des raisons de sécurité, nous avons adopté une conception passive plutôt qu'une mise en œuvre active.

2 Les objectifs du système proposé

Dans le cadre de ce projet nous avons proposé un système de surveillance réseau à base darknet pour la cybersécurité qui assure les fonctionnalités suivantes :

- La surveillance en permanence d'un bloc d'adresse non utilisées ;
- La collecte du trafic réseau destiné à ce bloc d'adresse ;
- Le prétraitement des données collectées pour améliorer la qualité de ces données afin d'optimiser leur exploitation future ;
- L'analyse des données collectées pour l'identification des incidents de sécurité ;
- Visualisation et perception des évènements détectés par le système de monitoring.

3 Architecture générale du système :

Le système de monitoring proposé est composé d'un ensemble de modules s'exécutant en pipeline comme montré dans la figure 15.

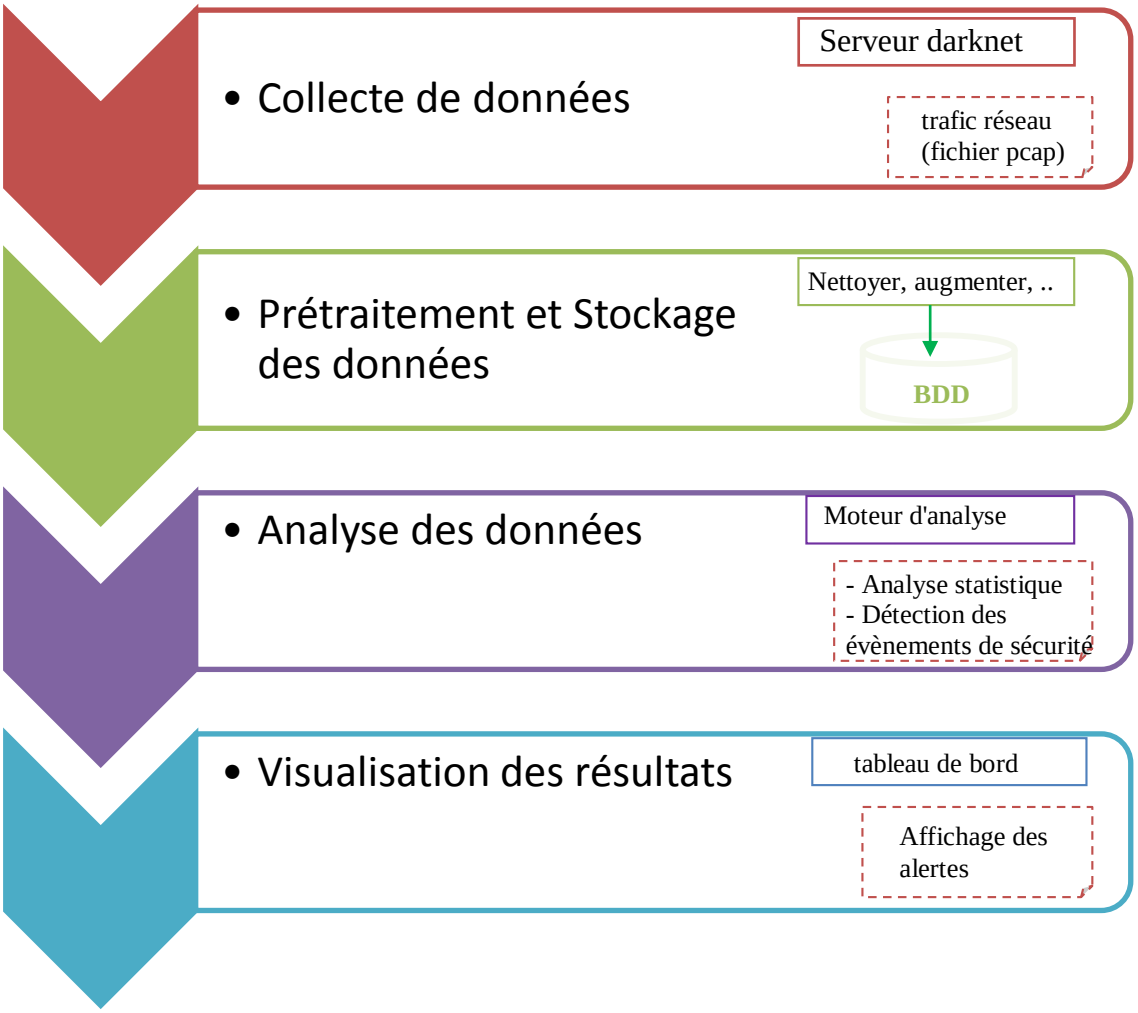


Figure 15. Architecture générale du système

- Collecte de données : il s'agit de surveiller et intercepter le trafic réseau destiné à un bloc d'adresse non utilisée (darknet). Pour ce faire, nous devons déployer un serveur darknet dans le réseau et e configuré pour que le trafic collecté ne soit contaminé avec du trafic ligitime.
- Prétraitement et stockage des données : les données collectées par le serveur darknet sont en format binaire (fichiers pcap); doivent être nettoyées (enlever les données

inutiles) et augmentées (par exemple, la géolocalisation ou les évènement coïncidant avec la période de collecte, ...).

- Analyse des données : une fois les données stockées dans la base de données elles seront prêtes à l'exploitation pour une analyse approfondie.
- Visualisation des résultats : Permet de visualiser les évènements de sécurité présentes sur le réseau détectées par l'analyse des données collectées.

4 Collecte de données

4.1 L'espace d'adressage utilisé :

Comme il est très difficile dans la première expérience de ce type de convaincre l'organisation de consacrer un grand nombre d'adresses IPv4, Le CERIST a mis à notre disposition un bloc de 33 adresses IP non utilisées. Cet espace est suffisant pour atteindre les objectifs du projet, mais aussi pour contribuer à des études récentes visant à évaluer l'efficacité du déploiement de ce type de systèmes dans de petits espaces [65].

4.2 Le déploiement du darknet

Du point de vue de la topologie du réseau, le capteur (serveur Darknet) est placé à l'extérieur du réseau de l'organisation sur la zone démilitarisée (DMZ) dans la partie publique du réseau.

Comme nous n'avons pas accès au routeur, les techniques de déploiement décrites dans la littérature s'avèrent inadéquates. Pour cela, nous avons essayé de trouver d'autres solutions pour capturer le trafic destiné à l'espace d'adressage alloué à l'étude.

Nous avons décidé de connecter le serveur au commutateur comme tout autre périphérique du réseau de recherche, avec des configurations spéciales en utilisant le pare-feu, et le filtrage du trafic pendant la collecte (figure 16).

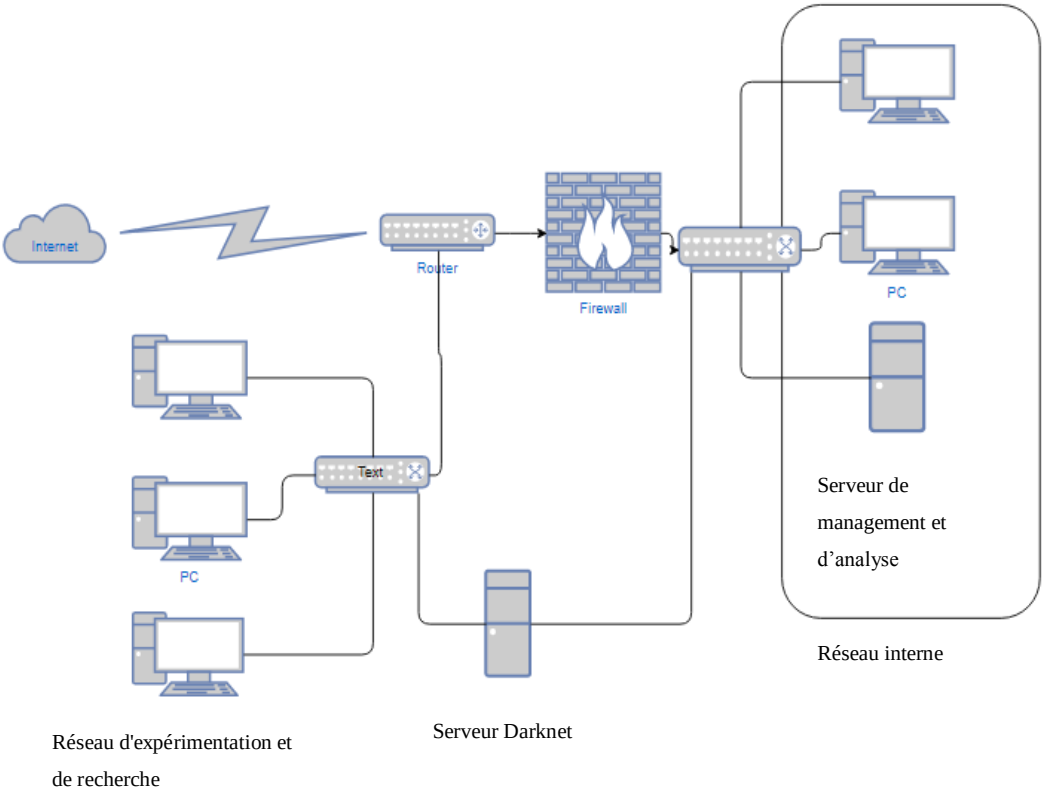


Figure 16. Le déploiement de serveur darknet

4.3 La configuration du serveur Darknet :

Pour le serveur Darknet, nous avons utilisé une machine avec un processeur i3 et 2 Go de RAM exécutant le système d'exploitation Ubuntu 18.04.

4.3.1 Configuration réseau :

Le serveur possède deux cartes réseaux :

- La première carte réseau est configurée avec toutes les 33 adresses IP allouées à l'étude (une adresse principale et les autres comme adresses IP secondaires[17]). Le but est de faire en sorte que la carte réseau réponde aux requêtes ARP correspondantes à ces adresses, par conséquent, le routeur dirige tous les paquets destinés à ces adresses vers le serveur darknet.

[17] Le protocole IP prend en charge les adresses secondaires.**RFC 791**, Section 3.2(internet protocol) :"le mappage entre les adresses d'hôte Internet et les interfaces réseau / hôte permettant à plusieurs adresses Internet de correspondre à une interface "

- La Deuxième carte réseau est configurée avec une adresse IP privée pour être accessible à partir de serveur management. Cela permet de ne pas contaminer le trafic darknet avec le trafic ligitime.
- Installer un serveur ssh(open ssh) pour sécuriser la connexion entre les deux serveurs.

4.3.2 Configuration du pare-feu du serveur darknet

Habituellement, l'attaquant envoie un ping sous forme d'un paquet ICMP echo-request, et attend une réponse sous forme d'un paquet ICMP echo-reply. Il peut aussi contacter un port TCP et attend un acquittement signifiant que la connexion est acceptée ou un **Reset** signalant une connexion est refusée ou "ICMP port unreachable", message indiquant que le port est injoignable (ex : le cas de paquet rejeté par le pare-feu).

Dans tous les cas, on comprend que la machine répond. Notre objectif est de garder l'espace d'adressage sombre pour que les adresses semblent inexploitées (Sans réponse). Pour ce faire, nous avons utilisé un pare-feu afin d'empêcher la machine de répondre.

Nous avons choisi **IPtable** comme pare-feu pour les raisons suivantes :

1. Nous pouvons gérer les règles pour chaque carte réseau séparément.
2. La fonction DROP : Quand un paquet atteint le pare-feu, la règle DROP peut Interdire à un paquet de passer et sans réponse.

Pour la première interface réseau (interface de collecte) : afin d'assurer la sécurité de serveur empêcher tous les paquets entrants et sortants sur les deux interfaces.

```
iptables -P INPUT DROP
iptables -P OUTPUT DROP
iptables -P FORWARD DROP
```

Pour la deuxième carte réseau (gestion) : Ouvrir uniquement le port 22 pour le serveur Openssh sur l'interface 2 et seulement pour l'adresse IP de serveur de management.

```
iptables -A INPUT -i eth1 -s 10.3.4.32 -p tcp --dport 22 -j ACCEPT
iptables -A OUTPUT -i eth1 -d10.3.4.32 -p tcp--sport 22-j ACCEPT
```

4.4 La capture du trafic darknet

Pour capturer le trafic destiné au bloc d'adresses surveillées, nous avons utilisé Wireshark qui est un logiciel de capture et d'analyse de paquets réseau le plus populaire. Il peut reconnaître plus de 2 000 protocoles contenant plus de 200 000 champs.

4.4.1 Les filtres de capture :

Notre objectif est de capturer le trafic provenant de l'extérieur du réseau, et comme mentionné précédemment, nous avons mis le serveur Darknet dans le réseau de production, donc nous avons utilisé les filtres de capture[18] pour masquer certains paquets de la liste de paquets capturé (les paquet arp et tous les paquets provenant de l'intérieur du réseau).

Les filtre de capture utilisés sont les suivants :

```
ether src @mac_router and not arp and not (src net @reseau_interne)
```

Le résultat de la capture est présenté par la figure 17.

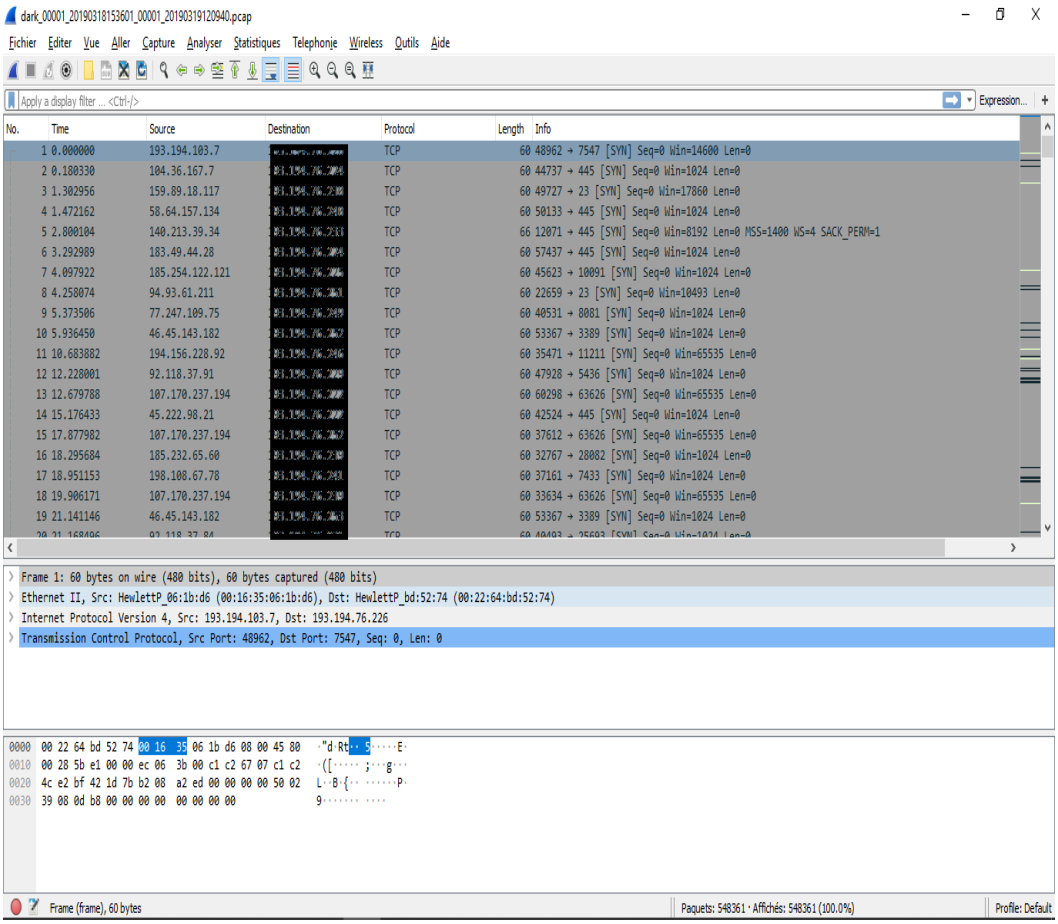


Figure 17.Résultats Wireshark, capture de paquets

4.4.2 Enregistrement des paquets capturés:

Nous avons configuré Wireshark pour stocker les paquets capturés dans des fichiers PCAP, en précisant la taille de chaque fichier à 50 Mo. Wireshark crée automatiquement un nouveau fichier lorsque la taille du fichier actuel atteint 50 Mo. Ce choix visait à assurer la stabilité

[18] Les filtres de capture sont définis avant de commencer une capture de paquet et ne peuvent pas être modifiés pendant la capture.

des performances du capteur(serveur). Wireshark a besoin d'environ 1Go de mémoire vive pour ouvrir un fichier PCAP de 50 Mo, qui est la moitié de la mémoire disponible sur le serveur.

Nous avons configuré wireshark pour utiliser un tampon circulaire avec 1000 fichiers (50Go), Wireshark suprimme automatiquement le premier fichier si le nombre total de fichier atteint 1000.

4.5 La configuration du serveur de management et d'analyse :

Pour le serveur d'analyse nous avons utilisé une machine avec un processeur i3 et 8 Go de RAM exécuté Ubuntu 18.04, avec et une capacité de stockage primaire 1To, évolutive dans le futur. Le serveur a une carte réseau configuré avec adresse IP de même sous réseaux privé Nous utilisons également le pare-feu IPtables avec une configuration simple pour bloquer tout le trafic entrant et sortant.

```
iptables -P INPUT DROP
iptables -P OUTPUT DROP
iptables -P FORWARD DROP
```

Sauf le trafic provenant du serveur darknet (avec le port source 22 ssh).

```
iptables -A INPUT -s 10.3.4.33 -p tcp --sport 22 -j ACCEPT
```

Le trafic vers le serveur darknet (avec port de destination 22 /ssh).

```
iptables -A OUTPUT -d10.3.4.33 -p tcp --dport22 -j ACCEPT
```

Nous avons configuré le serveur d'analyse pour copier automatiquement chaque nouveau fichier créé par Wireshark sur le serveur darknet en utilisant le logiciel Rsync[19] qui peut accéder aux fichiers et les déplacer vers le serveur d'analyse par connexion SSH.

```
rsync -v -e ssh
darkuser@10.3.4.32:/home/darkuser/capture/*.*/home/localuser/capture
```

[19]Rsync est un logiciel de synchronisation de fichiers en ligne de commande qui fonctionne de manière unidirectionnelle, il copie donc les fichiers d'un endroit vers un autre en synchronisant en fonction de la source (nous pouvons synchroniser ou copier les nouveaux fichiers créés).

5 Prétraitement et stockage des données :

5.1 Prétraitement des données :

Le problème auquel nous sommes confrontés lorsque nous traitons des fichiers PCAP est la quantité d'informations contenue dans ces fichiers. Chaque paquet contient plusieurs champs, dont Wireshark reconnaît environ 200 000 champs individuels.

Wirehark est l'un des meilleurs outils permettant de lire et d'analyser des fichiers pcap. Les données capturées par le darknet et stockées dans des fichiers pcap peuvent atteindre des centaines de gigas à long terme. Comme nous l'avons mentionné précédemment comme exemple, Wirehark avait besoin de 1 Go de RAM pour un fichier PCAP de 50 Mo. De plus les analyses possibles par Wireshark sont limitées, donc son utilisation pour analyser une grande quantité de données est contraignante.

Dans la plupart des cas, nous n'avons pas besoin de tous les champs dans l'analyse, c'est pourquoi nous devons extraire les informations utiles pour l'analyse à partir de fichiers PCAP et les mettre sous une autre forme pour les exploiter.

Utiliser une base de données pour stocker ces informations est l'une des solutions proposées, dans notre projet nous avons utilisé la base de données NoSQL Elasticsearch.

Pour que nous puissions entrer les données dans cette base de données, nous devons les convertir au format JSON à partir de fichiers pcap, et pour cela nous avons utilisé l'utilitaire Tshark

Tshark peut générer différents formats entre autres le format JSON pour l'API Elasticsearch Bulk.

```
tshark -r dark.pcap -T ek>dark.json
```

5.2 La création du modèle (template) :

Le fichier JSON généré par tshark contient toutes les données du fichier PCAP en format texte (chaine de caractères). Sans les bons types de données (Adresses IP, timestamp, numéro de port) nous ne pouvons pas effectuer les analyses adéquates. De plus le grand nombre de champs (comme mentionné ci-dessus, Wireshark connaît environ 200 000 champs par paquet), peut ralentir l'indexation et la vitesse de la requête sur elasticsearch.

Pour indexer les champs par leur type réel et pour n'indexer que les champs dont nous avons besoin, nous avons explicitement spécifié un mappage Elasticsearch.

Nous avons créé le Template (Voir l'annexe) et l'avons mis dans elasticsearch comme modèle de base pour l'indexation des paquets, le Template contient des commentaires pour définir le rôle des champs et pourquoi nous les avons choisis dans l'indexation.

5.3 Enrichissement des données :

Nous avons ajouté au Template un champg geoip de typeGeo-**point datatype** qui permet de stocker des points géographiques en longitude et en latitude de l'adresse IP source de chaque paquet.

Dans Elasticsearch, il y a le concept IngestPipeline. Un pipeline est la définition d'une série de processeurs pouvant apporter de nombreuses modifications aux données.

Nous avons ajouté une nouvelle pipline à elasticsearch contenant le processeur geoip (Voir l'annexe). Le processeur geoip ajoute des informations sur l'emplacement géographique des adresses IP, en fonction des données des bases de données Maxmind[66]. Ce processeur ajoute cette information par défaut sous le champ geoip.

5.4 Préparation Importation des données à partir des fichiers Json :

Afin d'automatiser l'ingestion des données collectées vers elasticsearch, nous avons utilisé l'outil Logstash. Ce dernier peut rechercher dans un répertoire de nouveaux fichiers(input), les traiter automatiquement et effectuer des transformations et des enrichissements plus complexes des données (filter) avant de les envoyer à Elasticsearch(output).

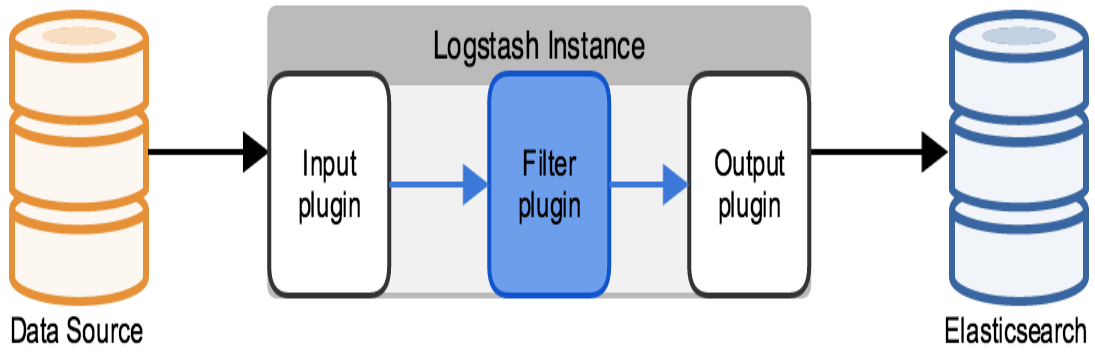


Figure 18.L'architecture de logstash

- Input : Pour l'entrée (input) nous avons installé le plugin logstash-input-file,ce qui permet de lire ligne par ligne les fichiers Json générés par tshark, dans le fichier de configuration nous avons spécifié le chemin de stockage de fichiers json pour que logstash surveille et traite les nouveaux enregistrements.
- Le traitement (filtre) : Pour le traitement de donnée nous avons installé et utilisé les filtres suivants :

  Json-filtre : nous permet d'extraire et de gérer la structure de données JSON, nous avons l'utilisé pour supprimer le champ message dans chaque enregistrement de paquet sur le fichier Json.

  Date-filtre : Logstash définira automatiquement un timestamp basé sur l'heure d'entrée(input) pour chaque enregistrement, nous avons utilisé date-filtre pour définir le temps de champ @timestamp pour lequel tous les paquets sont triés.

  Grok-filter : Le filtre grok extrait le nom de protocole réseau le plus interne du champ frame_frame_protocols (dont le format est « protocole : protocole : protocole », par exemple « eth:ethertype: ip: tcp: http») dans un champ «protocole» de niveau supérieur.

  - Output :

  Nous avons défini elasticsearch comme output en spécifiant l'adresse et le port du serveur elasticsearch. Nous avons spécifié aussi le pipeline qui doit être exécuté dans elasticsearch (geoip pipeline que nous avons déjà ajouté). Enfin nous avons ajouté les indexes avec le mot packet et la date de capture, donc chaque index avec le nom packets-`YYYY-MM-dd` contiendra les paquets de cette date.

Voir l'annexe pour les commandes d'installation des plugins, et le fichier de configuration de logstash (packet.conf).

Après l'exécution de Logstash les données sont ingérées dans elasticsearch selon le template créé.

6 Analyse des données

Nous avons effectué plusieurs analyses de données.Dans cette section, nous mentionnons les outils utilisés, et l'approche suivie dans chaque type d'analyse, sans mentionner les résultats, qui seront discutés en détail dans le prochain chapitre.

6.1 Analyse préliminaire :

L'objectif est d'effectuer une analyse de gravité basée sur les menaces en utilisant des IDS open source, Snort et Bro, deux NIDS open source, combinant les avantages de la signature, du protocole et de l'analyse d'anomalies, ont été mis en œuvre et utilisés.

Pour effectuer l'analyse des menaces, nous avons installé et configuré les NIDS dans le serveur de gestion et d'analyse. Nous avons analysé les fichier PCAP et enregistré les fichiers logs généré par ces IDS dans le même serveur pour futur utilisation et analyse.

6.2 Profiling de donnés darknet :

Nous avons utilisé kibana (voir section 7) pour la visualisation de donnés et ainsi pour l'analyse de donnés, avec la capacité de recherche fournie par elasticsearch et les visualisations pouvant être créés par kibana et en utilisant des requêtes spécifiques, nous avons extrait beaucoup d'informations sur la qualité du trafic, ses sources et les menaces potentielles et les comparons avec des travaux antérieurs.

6.3 Analyses approfondies :

la quantité énorme du trafic capturé à moyen et long terme nous oblige à utiliser des algorithmes de data mining et de machine learning pour extraire des informations utiles, malgré les nombreux programmes existe pour l'application de ces algorithmes, nous avons décidé de programmer notre propre application pour utiliser ces algorithmes afin de pouvoir les adapter à nos besoins qui correspondent à notre approche proposée.

6.3.1 Approche proposée :

Nous essayons de classer les adresses IP sources en plusieurs catégories en fonction du comportement des sources de trafic (IP source), en appliquant des algorithme de data mining et machine learning non supervisées sur le trafic extrait, nous décrivons des grappes d'activités partageant des comportements similaires.

Nous affirmons que l'analyse de ces grappes peut révéler les caractéristiques des comportements suspects, des programmes malveillants inconnus, cela permet de prendre des mesures préventives à l'avance contre les menaces inconnu (Zero Day Attacks, Forever-day vulnérabilités), et également de détecter le changement de comportement de certains logiciels malveillants connus.

Des travaux similaire utilise cette technique mais pour le clustering de paquets avec un nombre prédéfini de clusters [81,82], dans [83] Bou-Harb et al utilise k-means pour la classification de sources de scan sur le port 0 , mais toujours avec un nombre prédéfini de clusters et l'utilisation de fonction de nombreuses caractéristiques de paquets ,Nous pensons que la majorité de ces caractéristiques peuvent être manipulées et modifiées par une simple ligne de code dans les programmes malveillants(Ex :Tcp.windowsSize,TTl) ,donc leur valeur est presque aléatoire.

Pour cela nous avons inspirés de l'application de cette technique dans le domaine de classification de documents, dans ce dernier il s'agit d'une petite statistique concernant les mots, on enregistre le nombre d'occurrence et on le compare par rapport à la taille du document puis on déduit la fréquence, de la même manière, nous enregistrons le nombre de fois que chaque port TCP est ciblé par l'adresse IP et on le compare par rapport à le nombre de ports TCP ciblés puis on déduit la fréquence.

Note : Toutes les équations et tous les algorithmes mentionnés dans cette section peuvent être consultés dans le livre [80] .

6.3.2 Matrice adresse IP Port (MAP) :

Dans le domaine de classification de documents, il s'agit d'une matrice document terme, pour notre solution nous avons construire une matrice adresse IP/ Port à partir des données de paquets dans la base de données Elasticsearch, contient la fréquence de ciblage des ports par adresse, le tableau 5 montre un exemple.

La fréquence PFi= NP/N

PFi : (Port frequency) la fréquence de ciblage de port i par la même adresse.

N : Nombre de ports ciblés

NP : le nombre de fois que le port est ciblé par l'adresse

Tableau 5.Matrice adresse IP Port

@ IP/Port	23	2323	22	139	5555
87.251.81.86	4,60	0.33	0	0	1 ,34
185.153.197.61	0 ,012	0	5,60	0.33	0

6.3.3 Pondération PF-IPF :

Dans la classification de documents le TF-IDF (de l'anglais term frequency-inverse document frequency) est une méthode de pondération souvent utilisée en recherche d'information et en particulier dans la fouille de textes. Cette mesure statistique permet d'évaluer l'importance d'un terme contenu dans un document, relativement à une collection ou un corpus. Le poids augmente proportionnellement au nombre d'occurrences du mot dans le document. Il varie également en fonction de la fréquence du mot dans le corpus.

Pour la classification des adresses IP, Nous croyons que cette méthode aide à mettre en évidence un comportement de scanning inhabituel, les ports ciblés rarement auront une plus grande valeur (importance) permettant la plus grande distance entre l'adresse qui cible ce port et le reste des adresses.

$$IAFi = \log (A/N)$$

IAF : Fréquence inverse d'adresse

A : nombre total d'adresses dans le corpus.

N : nombre d'adresses où le Port i est ciblé.

$$PF\text{-}IPF(i) = PFi * IAFi$$

La valeur PF pour chaque port dans la matrice est multipliée par la valeur IAF appropriée.

6.3.4 Algorithme Implémenté :

- K-means : est un algorithme d'exploration de données et d'apprentissage automatique, qui prend des données en entrée et les regroupe en k clusters

connexes sans aucune connaissance préalable de ces relations, il utilise une mesure de similarité pour comparer les données, il fonctionne comme suit :

1. L'algorithme sélectionne arbitrairement k points comme centres de cluster initiaux (les moyens).
2. Chaque point du jeu de données est attribué au cluster fermé, en fonction de la distance euclidienne entre chaque point et chaque centre de cluster.
3. Chaque centre de cluster est recalculé en tant que moyenne des points de ce cluster.
4. Répétez les étapes 2 et 3 jusqu'à ce que, la condition d'arrêt (nombre d'itérations ou autre condition).

6.3.5 Mesure de Distance(similarité) Implémenté :

- Distance euclidienne : Cas particulier de Minkowski distance Sa Formule :

$$d(i,j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \cdots + |x_{il} - x_{jl}|^2}$$

d : la distance entre deux vecteurs i et j(une ligne de la matrice pour une adresse IP), x dans notre cas est le tf-idf d'un port.

- Cosinus similarité :

$$cos(d_1, d_2) = \frac{d_1 \bullet d_2}{\| d_1 \| \times \| d_2 \|}$$

Cos (d1, d2) : la distance entre deux vecteurs 1et 2, où • indique un produit vectoriel,et|| d || est la longueur du vecteur d.

6.3.6 Mesure de Qualité de clustering Implémenté (l'indice de bouldin) :

L'idée de cet indice est de comparer les distances intra-cluster (c'est l'homogénéité), que l'on veut faibles, aux distances inter-cluster (la séparation), que l'on veut grandes.

$$DB = \frac{1}{n} \sum_{i=1}^{n} \max_{j \neq i} \left\{ \frac{I(c_i) + I(c_j)}{I(c_i, c_j)} \right\}$$

I(Ci) représente la moyenne des distances entre les documents appartenant à la classe

(cluster) *Ci* et Son centre. Et *I(Ci,Cj)* représente la distance entre les centres des deux classes *Ci* et *Cj* . la meilleure partition est celle qui minimise la similarité entre les classes, et qui maximise la similarité dans les classes.
Note : l'indice de bouldin et l'algorithme de clustering doivent utiliser la même méthode de calcul de la distance.

6.3.7 Langage de programmation et technologies utilisées :

L'application est écrite en java nous avons utilisé la bibliothèque JavaFx[20] pour l'interface graphique, et pour faciliter l'interaction avec la base de données Elasticsearch nous avons utilisé la technologie JPA[21] (Java Persistence API), cette dernière traite uniquement des bases de données relationnelles, et pour cela nous avons utilisé le pilote JDBC Elasticsearch[22] pour convertir des requêtes SQL en requêtes Lucene.

Nous avons implémenté dans l'application et écrit à partir de zéro l'algorithme k-means, et les 2 mesures de similarités, cosine similarity et euclidean distance,

6.3.8 L'application :

Dans l'application, nous pouvons utiliser le graphique en courbes pour surveiller et évaluer le clustering pour chaque itération en fonction de l'indice de bouldin(figure 20), mais aussi pour et comparer et trouvez le meilleur nombre de grappes(clusters) K (voir le chapitre 5). Nous pouvons aussi également voir le résultat final de clustering (figure 21), ces résultats peuvent être envoyés à Elasticsearch en prenant chaque adresse du cluster, et en recherchant tous ses paquets et en les écrivant dans un fichier CSV,ce fichier est lu par logstash et est envoyé à Elasticsearch après le traitement nécessaire (identique à l'opération précédente avec json,il faut aussi utilisé un Template pour ces données)(figure 19) .

Ensuite, nous pouvons examiner les paquets de chaque Cluster et les analyser avec KIBANA.

[20] JavaFX est un framework et une bibliothèque, qui permet de créer une interface graphique pour des applications de bureau, des applications internet riches et des applications smartphones, permet aussi d'utiliser CSS pour améliorer l'apparence de l'application.
[21] La technologie JPA (Java Persistence API) a pour objectif d'offrir un modèle d'ORM (Object Relational Mapping), donc permet le traitement de chaque enregistrement dans la base de données comme un objet.
[22] https://www.cdata.com/drivers/elasticsearch/jdbc/. Une des solutions possibles.

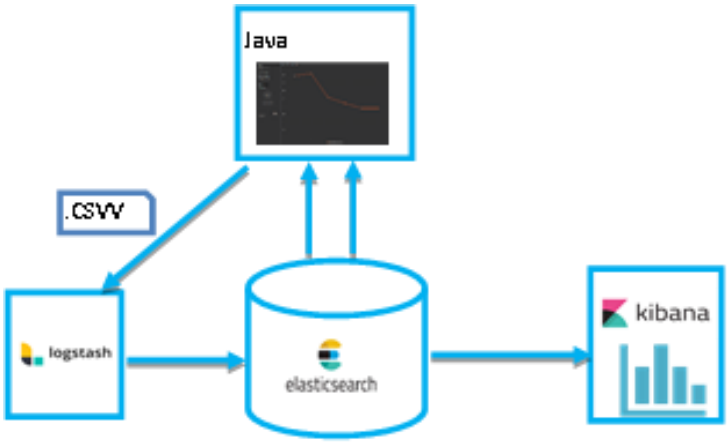


Figure 19.L'utilisation de l'application avec ELK

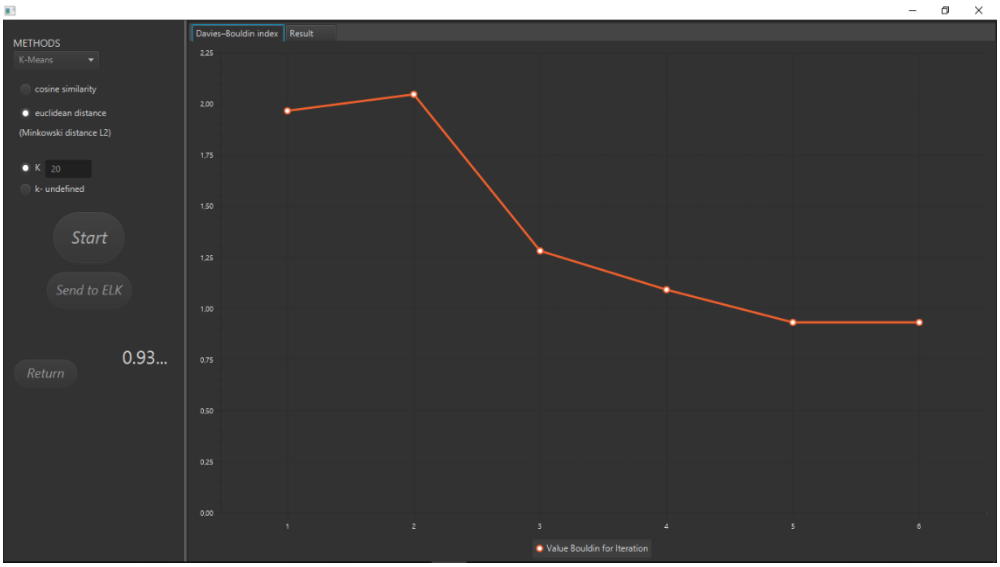


Figure 20.le resultat de clustering(Valeur Bouldin)

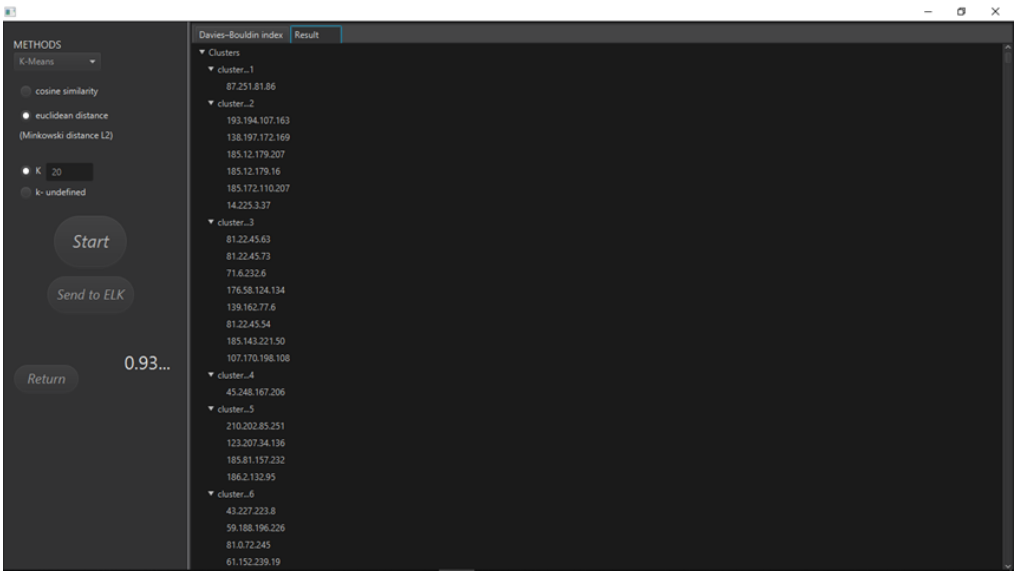


Figure 21.le resultat de clustering(les clusters)

7 Visualisation des données

Il existe plusieurs dashbords qui permettent la visualisation des données stockées dans elasticsearch. Nous avons choisi Kibana (la pile ELK) qui offre des tableaux de bord intéressants pour la visualisation ainsi que des consoles d'exécution de requêtes sur elasticsearch, allant des requêtes simples aux requêtes complexes, qui respectent la syntaxe de requête de Lucene[23].

Kibana permet de créer des visualisations sous forme de graphiques, de tableaux et de diagrammes, aidant ainsi à visualiser toutes les données qui ont été stockées dans Elasticsearch facilement. En créant des visualisations et utilisant les requêtes sur kibana pour filtrage et la recherche, nous pouvons facilement donner un sens aux données et obtenir des réponses aux questions qu'il aurait pu se poser pendant le processus d'analyse des données. Ces visualisations construites peuvent être utilisées pour la construction de tableaux de bord sur KIBANA (Dashboard).

La page suivante (la figure 22) sur kibana, montre les enregistrements dans stockés sur elasticsearch.

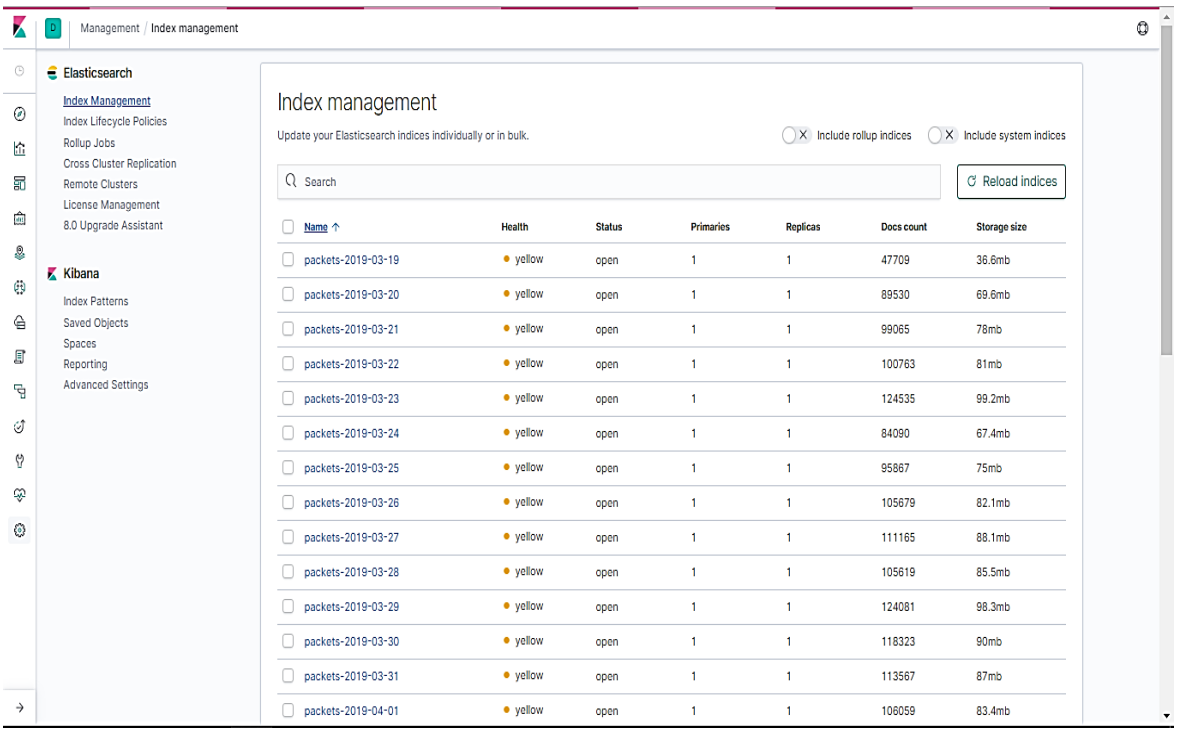


Figure 22.Les fichiers indexés dans elasticserach

[23]Lucene est une bibliothèque open source écrite en Java qui permet d'indexer et de chercher du texte. Il est utilisé dans certains moteurs de recherche.

Avant de pouvoir commencer à utiliser les données et à créer des visualisations pour l'analyse de données, il faut configurer l'Index patterns dans kibana. Les Index patterns sont utilisés pour identifier l'index Elasticsearch auquel la recherche et l'analyse s'exécutent.

Elasticsearch propose généralement deux types d'index :

- Index de séries chronologiques : Lorsque ces données sont stockées dans Elasticsearch, elles sont stockées dans plusieurs index (index roulants), les noms des index étant généralement ajoutés par un horodatage (ex Packets-2019-03-20, Packets-2019-03-21) .
- Index normaux : si les données ne contiennent pas d'horodatage et que les données n'ont pas de corrélation avec le temps, elles sont alors appelées données régulières. En règle générale, ces données sont stockées dans des index simples. Par exemple, données de catalogue de produits.

Dans notre cas, nous avons utilisé Index de séries chronologiques avec le Template Packets-* que nous avions créé dans elasticsearch.

Sur la page Discover de Kibana nous pouvons explorer les données et effectuer de manière interactive des requêtes de recherche, de filtrer les résultats de la recherche et d'afficher les données.

A ce niveau-là, les données darknet collectées sont prêtes pour être exploitées dans des analyses approfondies.

8 Schéma fonctionnel du système de monitoring darknet

En résumé la figure 23 montre la structure détaillée du système proposé et les outils utilisés. Le serveur darknet collecte le trafic réseau au moyen du logiciel Wirehark et les stocke dans des fichiers pcap.

En raison du volume de trafic entrant attendu à long terme ces fichiers sont déplacés sur une autre machine (le serveur d'analyse) pour le traitement et l'analyse.

Tshark génère d'autres fichiers json à partir de ces fichiers pcap. Les fichiers Json sont traités par logstash pour ajouter des données ou modifier la structure afin de les stocker dans la base de données NoSQL Elasticsearch.

Enfin, la base de données peut être utilisée pour l'analyse à l'aide de langages de programmation, et pour visualiser et analyser et explorer les données dans Kibana.

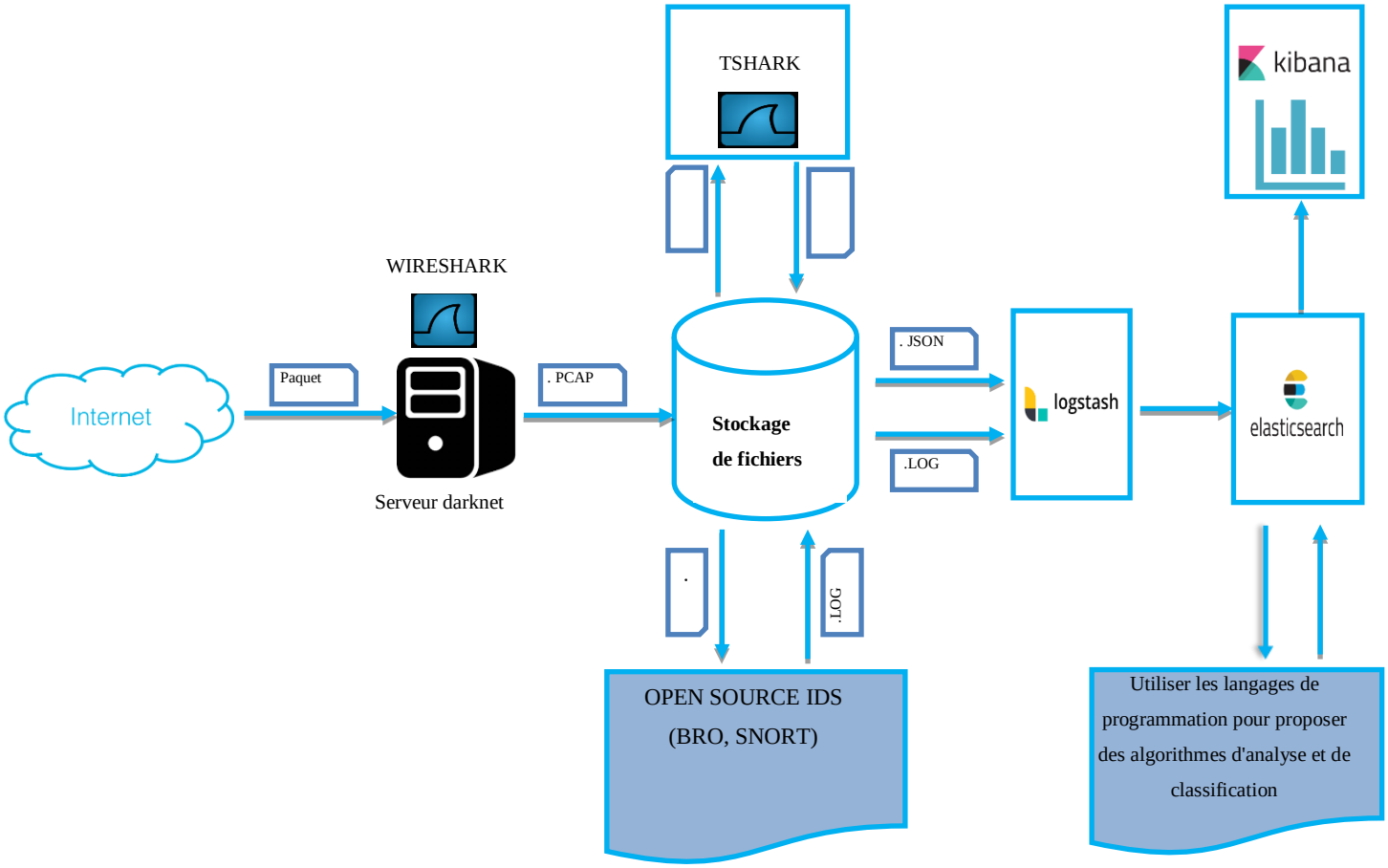


Figure 23.Schéma fonctionnel du système proposé

9 Conclusion

Nous avons présenté le système de monitoring que nous avons mis en place ainsi que les outils utilisés pour la collecte, le stockage et la visualisation des données darknet. Nous avons également présenté les prétraitements que nous avons apporté aux données en guise de préparation à l'exploitation dans des analyses approfondies.

Pour l'analyse nous avons présenté les outils utilisés pour l'analyse préliminaire (IDS), l'approche proposé pour une analyse approfondie, et l'application créée pour la mise en œuvre de cette approche.

Dans le chapitre suivant, nous allons présenter les résultats d'analyses des données darknet collectées par le système mis en place.

Chapitre 5 : Analyse et résultats

1 Introduction

Ce chapitre présente un aperçu des résultats extraits des données collectées par le serveur darknet, qui surveille un espace d'adressage de 33 adresses IP.

L'analyse porte sur les données de trafic collectées entre le 19Mars2019(12 :09) et le 21Avril2019(21 :06). La taille combinée des fichiers PCAP du trafic collecté est d'environ 391 Mo, et les fichiers JSON résultant de ces fichiers d'environ 10,2 Go. Un total de 3726342 paquets IPv4 provenant de **238268** adresses IP sources différentes envoyées à 33 adresses IP de destination différentes ont été collectées au cours de cette période.

Ce chapitre commence par décrire la composition du trafic. Ensuite, une analyse approfondie du trafic est effectuée.

2 Analyse de la nature du trafic :

2.1 La composition du trafic :

Pour mieux comprendre la nature du trafic darknet, son contenu et ses menaces, nous avons effectué un profilage du trafic darknet observé.

2.1.1 Distribution des protocoles :

L'analyse commence au niveau de la couche réseau (IP) de la pile TCP / IP, les protocoles sont enregistrés dans les datagrammes IPv4 sur le champ protocole, cette valeur binaire de 8 bits indique le type de données utiles transportées par le paquet, ce qui permet à la couche réseau de transmettre les données au protocole de couche supérieure approprié. Les valeurs habituelles sont notamment ICMP (1), TCP (6) et UDP (17), dans les données de elasticsearch le champ est indexé sous le nom (layers.ip.ip_ip_proto).

La figure 25 et le tableau 6 illustrent la répartition du trafic reçu par protocole, les résultats et la figure 26 illustre le volume de trafic capturé par jour et par protocole .On peut remarquer que les paquets TCP représentent plus de 90% du volume de trafic presque chaque jour. Le protocole UDP (User Datagram Protocol) et ICMP se partagent le reste avec des parties variables au cours de cette période.

Il existe d'autres protocoles moins représentés tels que le protocole de transmission de contrôle de flux (SCTP) 1634 paquets, l'encapsulation IPv6 26 paquets, tous envoyés à partir

de la même adresse source (131.193.34.220, USA) et l'encapsulation de routage générique (GRE) 21 paquets.

Il est important de noter que certains logiciels de scan tels que Nmap [67] implémentent le scan SCTP et IGMP dans sa boîte à outils.

La dominance de TCP peut s'expliquer par le fait que la majorité des attaques scanning utilisent TCP.

Nous observons que le nombre de paquets TCP augmente sur la figure 26, en particulier pendant la période allant du 18 avril à 07h00 au 20 avril à 00h00. Cela nécessite une analyse temporelle approfondie et une comparaison de ces phénomènes susceptibles de révéler et d'expliquer la survenue de certaines attaques à des périodes déterminées et leur absence pendant d'autres périodes.

Tableau 6.Nombre et pourcentage de paquets par protocole

	TCP	UDP	ICMP	Autre
Paquets	3387188	287751	49743	1660
Pourcentage	90.9%	7.72%	1.33%	0.05%

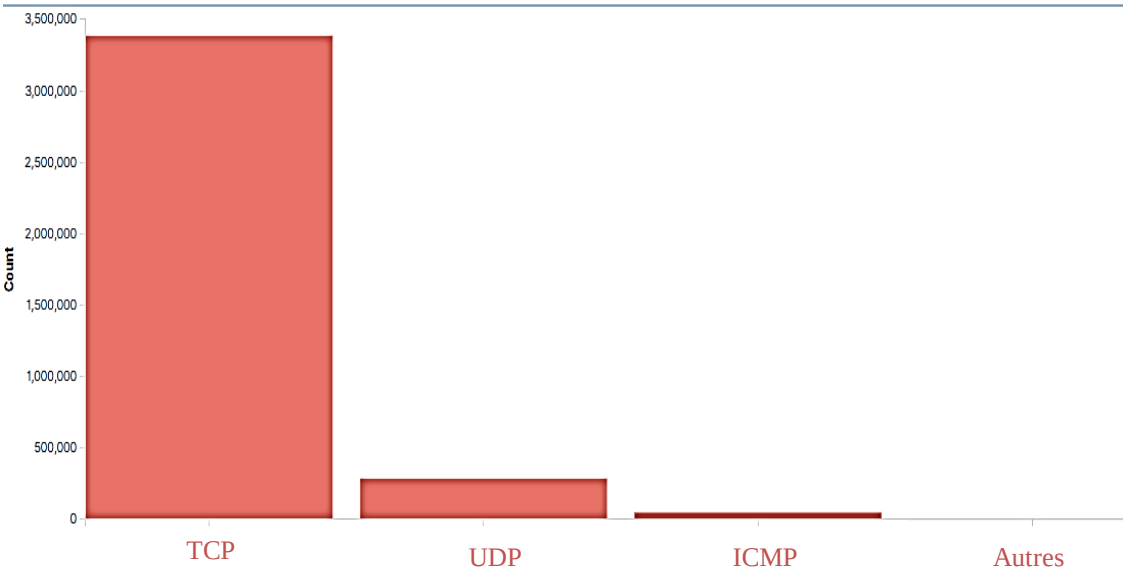


Figure 25Distribution de protocoles

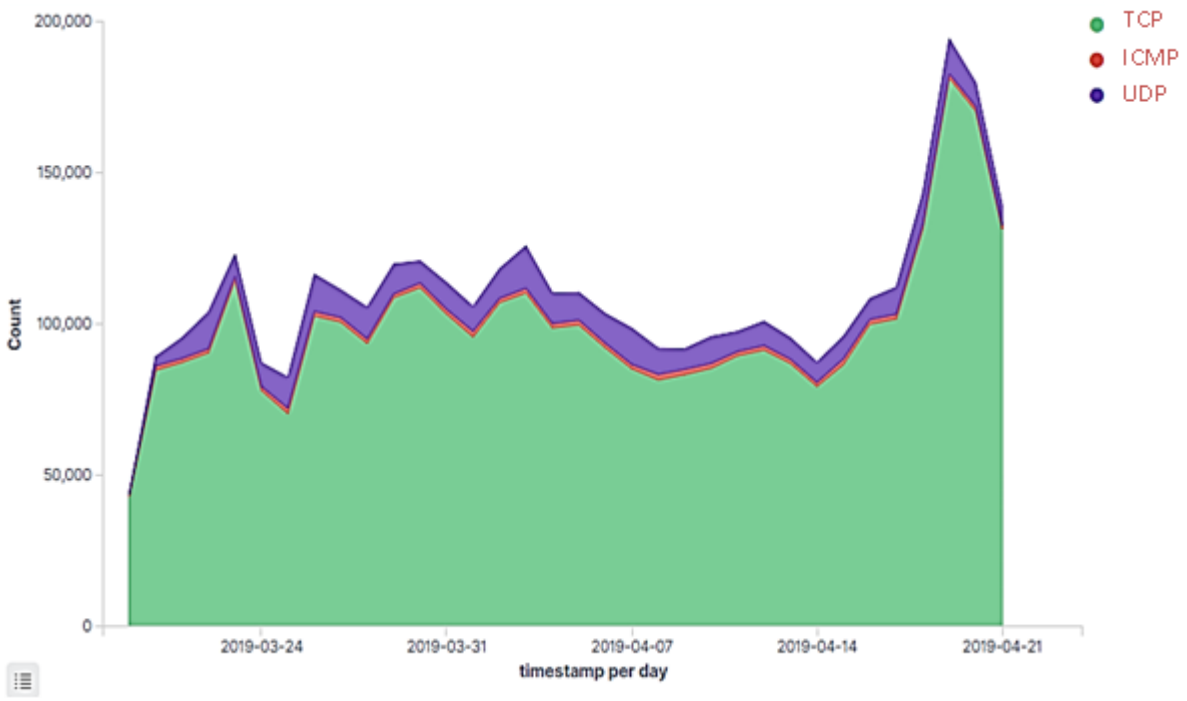


Figure 26.Volume de trafic quotidien par protocole

2.1.2 Les ports TCP :

Le trafic TCP représente la majeure partie du trafic observé, comme décrit précédemment, nous avons effectué une autre analyse sur le protocole TCP dans le trafic darknet collecté. Plus précisément, nous avons cherché à identifier les ports de destination. Ces informations pourraient révéler les ports ciblés utilisés dans les cyberattaques. Les figures 27 et 28 illustrent ces résultats.

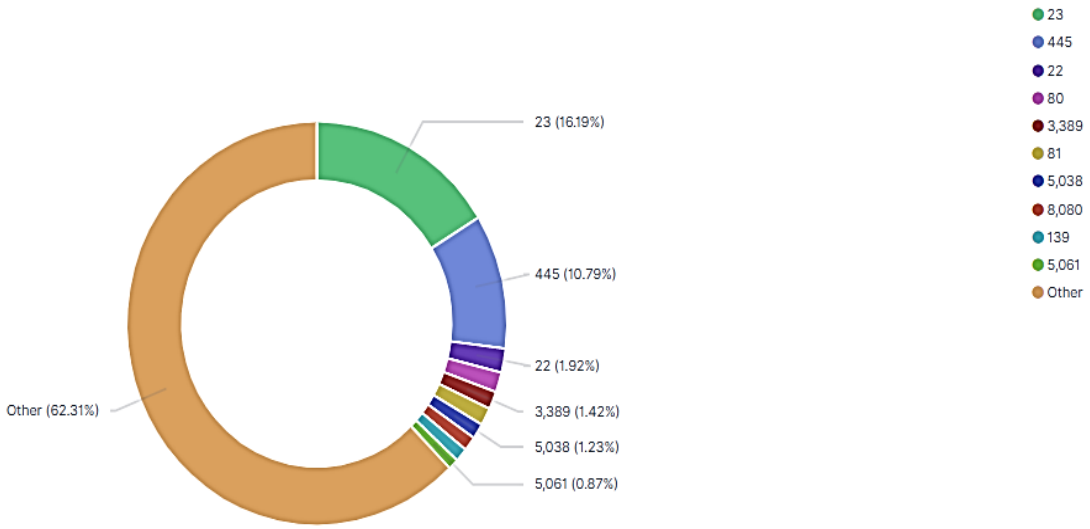


Figure 27.Top 10destination Ports TCP

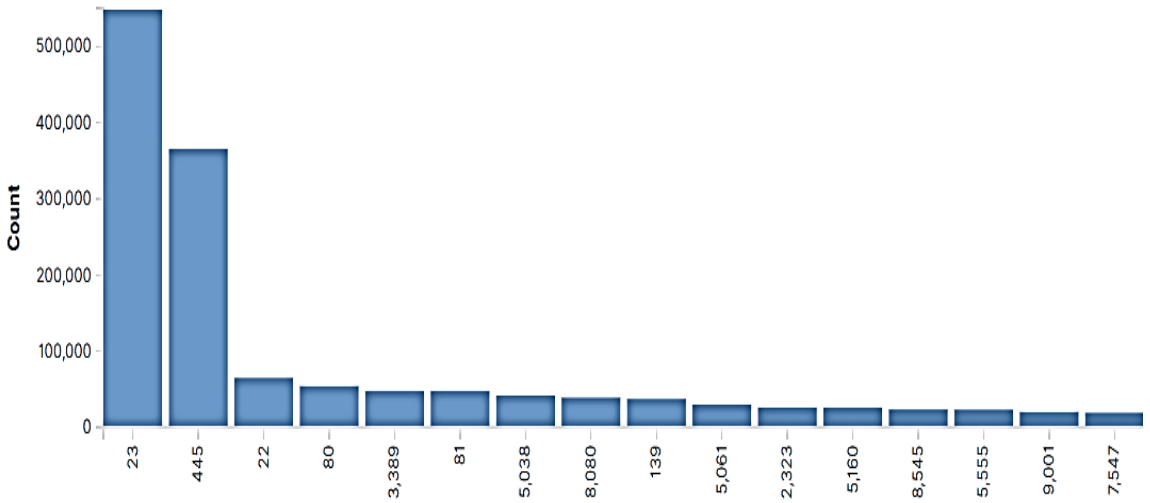


Figure 28.Principaux ports ciblés TCP

La figure 27 illustre les 10 ports TCP les plus ciblés pendant la période de collecte du trafic. Bien que l'on ait observé que le trafic était destiné à 22996 ports, le trafic ciblé sur ces 10 ports représente 37,69 % du trafic TCP total capturé.

A l'exception du port 5038 et le port 81, les autres ports sont tous des services bien connus attribués par l'IANA. Les services correspondants aux 5 premiers ports ont déjà souffert de problèmes de sécurité et de vulnérabilités [68,69,70,71]. Les appareils exécutant le service Telnet (port 23), Secure Shell (SSH /port 22) sont vulnérables aux attaques par Brute force [72], les port 445 et 3389 sont utilisé par les applications liées à Microsoft telles que le SMB/445 (Server Message Block), ou Microsoft Active Directory/445, Microsoft terminal server/3389, et ces services avaient une certaine vulnérabilité (Ex : SMBv1 [73]). Les serveurs Web sont sujets à des vulnérabilités telles que l'injection XSS et SQL. Le port 139/NetBIOS[24]et les ports d'application Web communs tels que TCP / 80 (et TCP / 8080 à un taux inférieur) sont également ciblés.

2.1.3 Les ports UDP :

Pour UDP nous avons aussi cherché à identifier les ports de destination. Les figures 29 et 30 illustrent ces résultats

[24]NetBIOS est utilisé principalement par Microsoft. C'est système permet d'établir des sessions entre différents ordinateurs d'un réseau. Développée par IBM

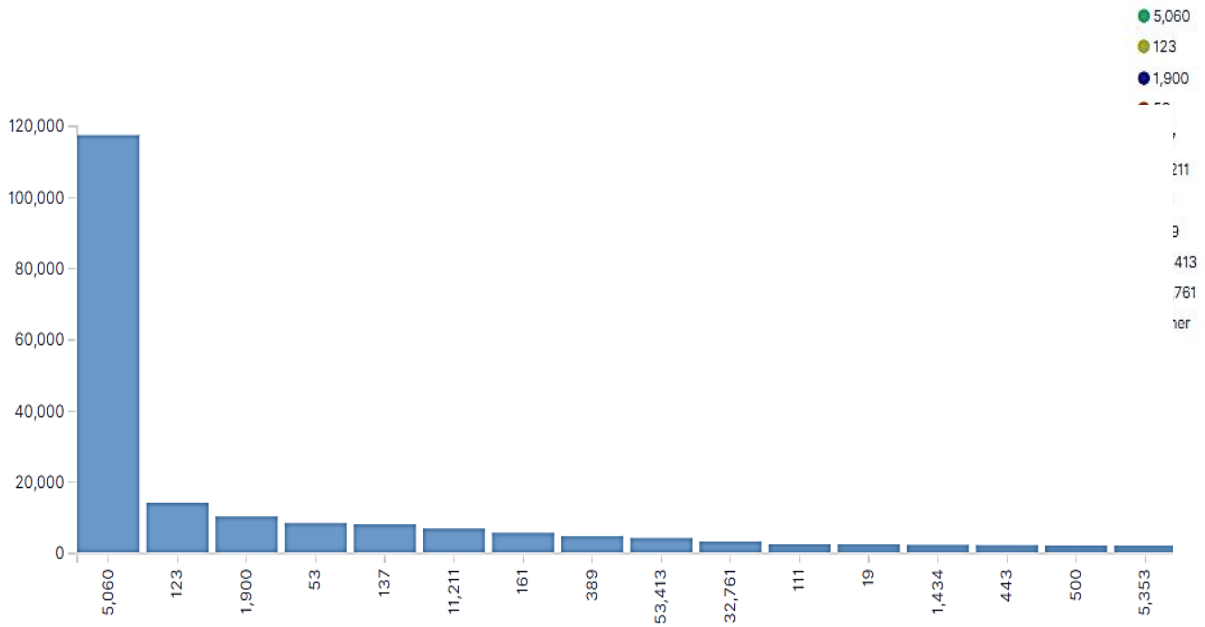


Figure 29.Top 10 destination Ports UDP

Figure 30.Principaux ports ciblés UDP

La figure 29 illustre les 10 ports UDP les plus ciblés pendant la période de collecte du trafic. Bien que l'on ait observé que le trafic était destiné à 1312 ports UDP, le trafic ciblé sur ces 10 ports représente 64.27% du trafic UDP total capturé.

Il semble clair que le port le plus ciblé est 5060, il représente 40.89%. Le port 5060 est le SIP[25], Le port 53413 est une porte dérobée connue dans les routeurs Netis[26]détectée pour la première fois en 2014. Toute charge utile envoyée au port UDP 53413 de ce périphérique est automatiquement exécutée [75]. Le port 123 NTP (Network Time Protocol) et le port 1900 pour le SSDP[27] (Simple Service Discovery Protocol) sont utilisés dans les attaques DRDOS [74], ou il s'agit très probablement de trafic généré par des réseaux de zombies à la recherche de machines avec ce port ouvert en prévision d'attaques DDoS futures. Le port 53, normalement dédié au service de noms de domaine (DNS) et le port 137 NetBIOS ont également été ports cibles. Une vulnérabilité sur le service NetBIOS affectant toutes les

[25]Session Initiation Protocol (SIP) est un protocole standard ouvert de gestion de sessions souvent utilisé dans les télécommunications multimédia le plus courant pour la téléphonie par internet (VoIP).
[26]Marque principalement présente en Chine, mais on peut la trouver dans d'autres régions.
[27] Le protocole SSDP (Simple Service Discovery Protocol) est un protocole réseau pour la publication et la découverte de services réseau.

versions de Microsoft Windows, permet à un attaquant de contourner le pare-feu ainsi que les périphériques de traduction d'adresses réseau (NAT) [76]. Le port 161 pour le protocole SNMP (Simple Network Management Protocol), est souvent utilisé à des fins de reconnaissance car SNMP est fréquemment installé sur des systèmes par défaut.

2.1.4 La répartition par nature du trafic TCP:

Le trafic Darknet TCP peut être classé en trois types distincts basés sur différentes causes fondamentales de ces activités, le scanning est en grande partie le résultat d'hôtes infectés sur Internet qui tentent de trouver d'autres cibles vulnérables, la rétrodiffusion (Backscatter) est le plus souvent le résultat d'attaques par déni de service(DDoS), et enfin une mauvaise configuration (Misconfiguration), qui résulte d'erreurs logicielles ou matérielles, nous avons classé les paquets du trafic en fonction de leur type selon la méthode décrite dans [28,77] .Nous classifions les paquets en fonction des combinaisons de TCP flags:

- Les paquets TCP SYN en tant que trafic de scan.
- Les paquets TCP SYN+ACK, RST, RST+ACK et ACK en tant que trafic de rétrodiffusion (Backscatter), car ces paquets sont susceptibles d'être générés par des hôtes qui tentent de répondre à une communication émise par une source usurpée avec une adresse dans le darknet.
- Le trafic restant comme une mauvaise configuration

Le tableau 7 présente les combinaisons des indicateurs (TCP flags) utilisés dans les paquets TCP collectés, et le nombre total de paquets et le pourcentage par rapport au trafic TCP pour chaque combinaison. Il est très clair que la domination de paquets ne contenant que le drapeau SYN avec plus 99.96%. Nous observons aussi l'absence de paquets contenant l'indicateur SYN-ACK. Cela indique une augmentation de l'activité de scanning par rapport aux résultats obtenus dans [28] (Dans ce travail également observé une augmentation de 63% en 2006 à 94% en 2010) et une absence de l'activité de DDoS basé sur la technique de SYN flood, ceci est également cohérent avec la tendance observée dans le même travail [28] (où les paquets ACK + SYN diminue considérablement de 26,1% en 2006 à 5,2% en 2010).

Tableau 7.les combinaisons des indicateurs (TCP flags) utilisés dans les paquets TCP

ACK	FIN	PSH	RST	SYN	URG	Paquets	Pourcentage
0	0	0	0	1	0	3385959	99,96%

0	0	0	1	0	0	688	0,020%
1	0	0	1	0	0	442	0,013%
1	0	1	0	0	0	4	0,00011%
1	1	0	0	0	0	95	0,0028%

Le tableau 8 illustre la distribution des paquets par nature du trafic TCP, les activités de scanning constituent la majorité du trafic darknet TCP.

Il est intéressant de noter que les résultats sont assez significatifs par rapport aux résultats obtenus par Claude et al [77] montré dans le tableau 10.

Il est intéressant aussi de noter que les résultats des travaux [77] (2012) et des travaux [28] (2010) (qui sont plus proches de nos résultats) sont très différents, ce qui renforce ce que nous avons noté précédemment que la surveillance d'un bloc d'adresses différent peut donner des informations différentes sur les menaces.

Il est également nécessaire de noter aussi que cette classification est considérée comme préliminaire, car les paquets classés comme Rétrodiffusion ne sont pas nécessairement le résultat d'attaques DDoS, car plusieurs de ces combinaisons peuvent être utilisées pour effectuer un scan de port [67]. Par exemple lorsque l'indicateur ACK est défini, un attaquant peut déterminer si un port est filtré ou non par un pare-feu et le pare-feu est avec ou sans état. De plus, les paquets RST peuvent être des attaques RST essayant de mettre fin à une connexion valide entre deux hôtes.

Tableau 8.La distribution de paquets par nature du trafic TCP

Scanning Traffic	Rétrodiffusion	Mauvaise configuration
99.96%	0.03%	0.01%

Tableau 9.La distribution de paquets par nature du trafic TCP[77]

Scanning Traffic	Rétrodiffusion	Mauvaise configuration
68.02%	2.00%	29.98%

Tableau 10..La distribution de paquets par nature du trafic TCP [28]

Scanning Traffic	Rétrodiffusion	Mauvaise configuration
93.9%	5.9%	0.2%

2.1.5 ICMP :

Il est le dernier des principaux groupes de protocoles. Chaque paquet de ce type contient deux éléments principaux : le type du paquet et un code qui peut fournir des détails supplémentaires [78]. Le tableau 11 montre les combinaisons de valeurs rencontrées lors de l'analyse.

Le trafic écho Type 8 constitue la majeure partie du trafic (96.76%), le Type 3 (Destination inaccessible) constitue (2.58%) et 11 (TTL dépassé) constitue (0.47%). Les types 3, 8 et 11 combinés constituent 99,81 % du trafic ICMP.

Tableau 11.ICMP types et codes

TYPE	CODE	Paquets	Pourcentage
8	0	48133	96,76336%
3	3	715	1,437388%
3	2	311	0,625214%
3	10	214	0,430211%
3	0	23	0,046238%
3	1	15	0,030155%
3	4	5	0,010052%
3	13	1	0,00201%
11	0	234	0,470418%
13	0	87	0,174899%
0	0	4	0,008041%
193	171	1	0,00201%

Les demandes d'écho ou de ping sont les paquets les plus fréquemment observés et sont couramment utilisés à la fois comme moyen de reconnaissance et pour détecter l'existence d'un hôte cible. De nombreux outils utilisent un ping pour déterminer si un hôte est joignable avant de commencer les scans sur les ports TCP et UDP. Certains outils tels que Nmap fournissent des options pour remplacer ce comportement par défaut et effectuer des analyses même dans le cas où les hôtes ne répondent pas.

Le type 0, sont des réponses d'écho qui sont générées en réponse à des paquets de type 8(demande d'écho) avec de fausses adresses sources, presque pareil pour le type 11 le message est envoyé pour informer l'expéditeur que le datagramme a été détruit lorsque le temps de vie d'un datagramme (code = 0) ou le temps de réassemblages des parties d'un datagramme (code = 1) est dépassé.

Le type 3 (Destination inaccessible) peut être un trafic de rétrodiffusion, sont générés par des routeurs ou des hôtes ou des systèmes de filtrage en réponse à des routes manquantes. Les valeurs de code généralement observées sont définies dans la RFC 792 [78], le code le plus enregistré dans le trafic darknet ICMP type 3 est le code 3 port inaccessible (port unreachable), ces paquets sont probablement générés par des systèmes de filtrage tels que des pare-feux.

Il existe un paquet avec valeur de type non défini 193, peut être généré à partir d'un périphérique mal configuré.

2.1.6 Tailles des paquets :

Le champ Longueur totale de 16 bits dans l'en-tête du paquet IPv4 indique la taille globale du paquet reçu, y compris l'en-tête et les données, en octets, sans la taille de l'en-tête de couche liaison. Sa valeur minimale est de 20 octets (un en-tête de couche IP de 20 octets et 0 octet de données) et sa valeur maximale est de 65 535 octets. Dans les données d'elasticsearch le champ est indexé sous le nom (layers.ip.ip_ip_len).

La figure 31 montre la longueur des paquets les plus enregistrés.

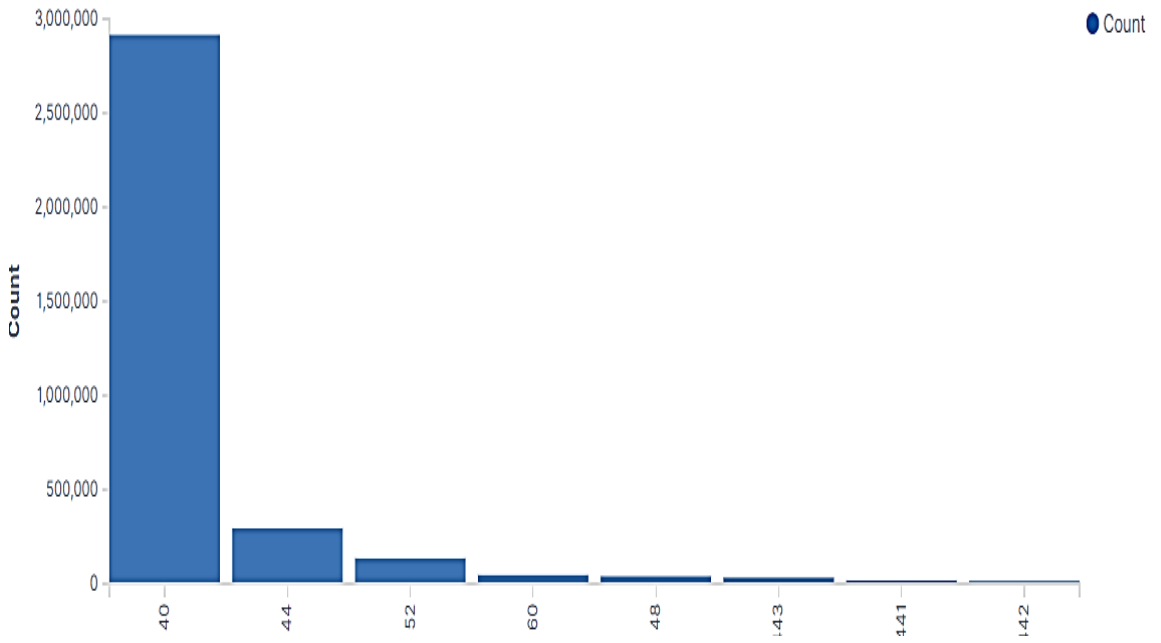


Figure 31.Longeur des paquets

Nous avons observé que plus de 94% des paquets ont moins de 60 octets. L'en-têtes TCP [79] a une longueur comprise entre 20 et 60 octets. Les en-têtes UDP [76] et ICMP [78] ont chacun 8 octets, donc la majeure partie du trafic observé est un scanning malveillant avec peu de paquets de données utiles. Cela confirme le résultat de la répartition par nature du trafic TCP dans la section 2.1.4.

La taille minimale observée est de 28 octets et la taille maximale est de 1500. Cinq paquets de taille maximale provenaient de la même adresse source le 15Avril2019 en visant le port UDP 53413. Cette adresse était probablement à la recherche d'un routeur Netis qui présentaient toujours la vulnérabilité de la porte dérobée mentionné dans la section 2.1.3.

3 Analyse et extraction des informations sur les menaces :

L'objectif de cette analyse est la découverte des menaces, la détection et la géo-localisation des sources de trafic malveillant ou d'anomalies, et de classer leurs sévérités. Cela permet d'avoir une vue d'ensemble qui aiderait aux procédures d'atténuation des menaces et de générer des listes noires catégorisées d'adresses IP qui se comportent de manière suspecte.

3.1 Distribution géographique :

La géo-localisation darknet nous permet d'identifier les sources du trafic darknet. Nous effectuons une géo-localisation pour déterminer les pays qui contribuent au trafic en fonction du volume de trafic et en fonction du nombre d'adresses IP.

La figure 32 illustre le HeatMapen fonction du volume de trafic provenant de chaque pays. Les pays sources ont atteint 200 pays où la majorité du trafic provenait de la Russie (38.3%), les Etats-Unis (24.7%) et la Chine (13.25%).

Il est intéressant de noter que dans notre analyse, la Russie occupe la première place contrairement aux travaux précèdent (l'Etats-Unis) [58,68].

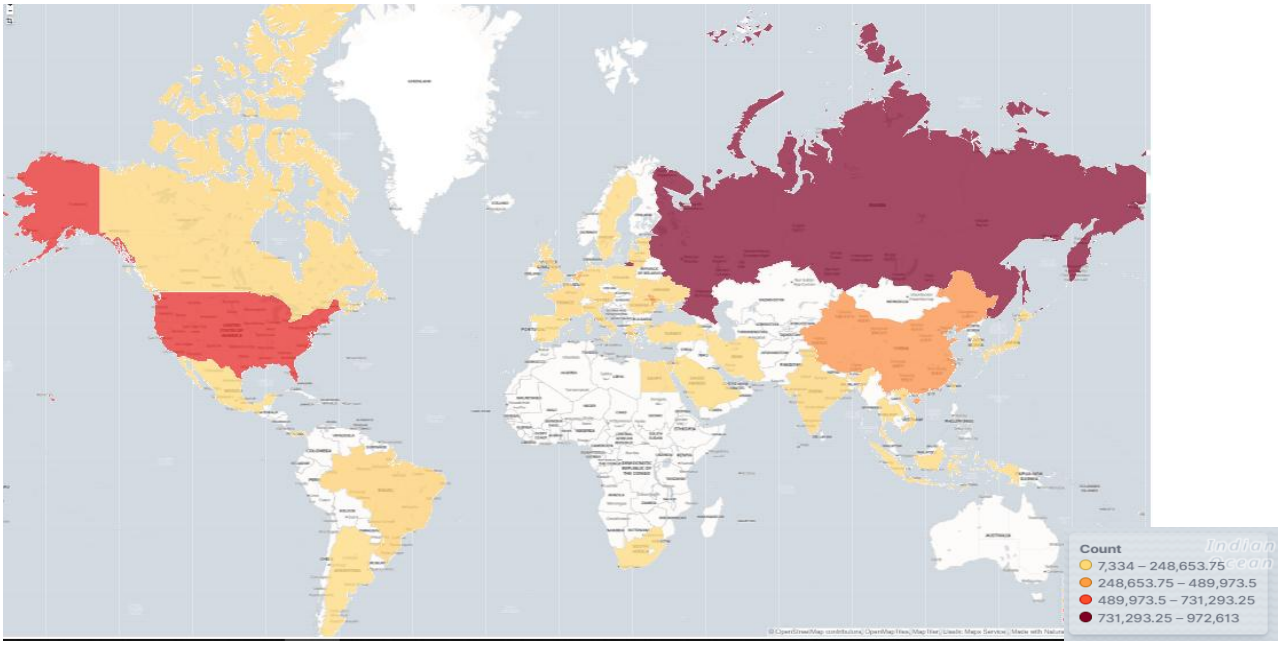


Figure 32. Localisation géographique en fonction du volume de trafic provenant de chaque pays

La figure 33 illustre le HeatMapen fonction du nombre d'adresses IP, la majorité des IP sources sont situées en Chine. Il est remarquable que l'Egypte et le Brésil, la Chine et la Russie représentent la majeure partie des d'adresses IP source par rapport aux autres pays.

Il est tout aussi remarquable que l'Égypte occupe la troisième source d'adresses IP, bien qu'elle n'apparaisse dans aucun des travaux précédents parmi les dix premiers, ce qui nécessite une enquête plus approfondie sur le trafic en provenance de ce pays.

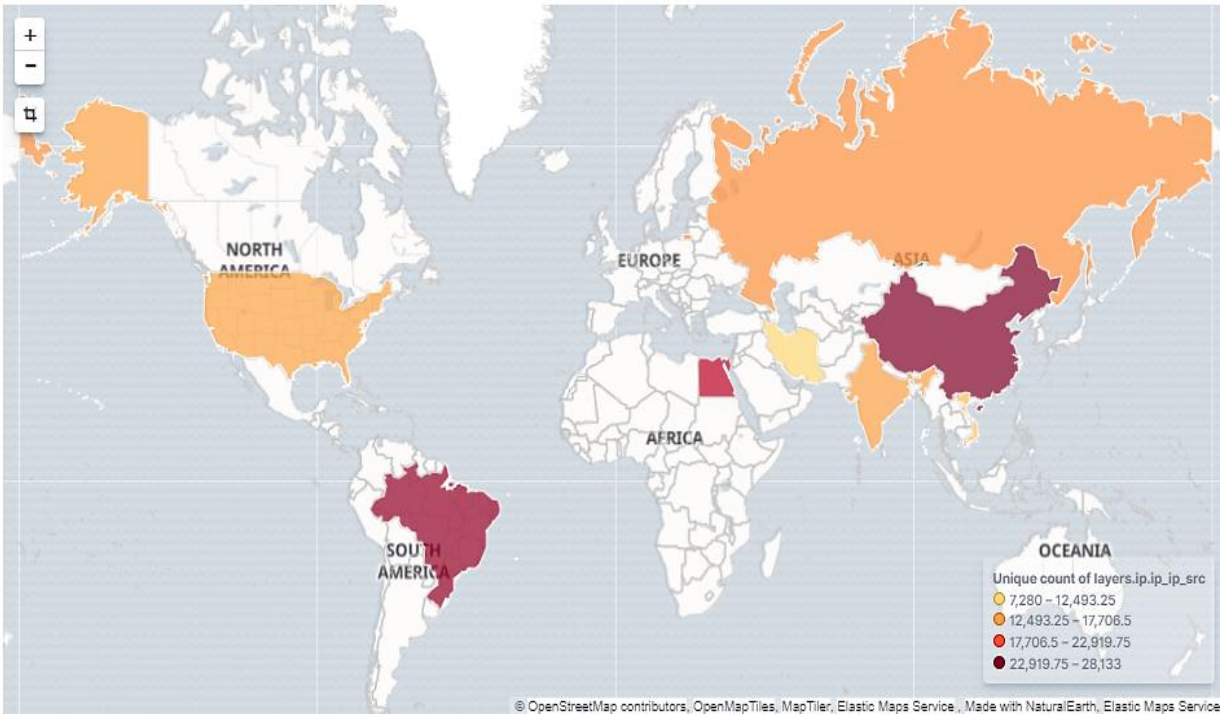


Figure 33.Localisation géographique en fonction du nombre d'adresses IP source

3.2 Analyse par NIDS :

Nous étendons notre tâche de profilage, en analysant les fichiers PCAP par l'utilisation de deux NIDS open source Snort et Bro, les résultats sont présentés ci-après.

3.3 DDoS NTP :

Les résultats de Snort révèlent 11 menaces de haute priorité de type DDoS NTP. Le protocole NTP (Network Time Protocol) est utilisé par les machines connectées à Internet pour synchroniser leurs horloges. Les anciennes versions de NTP prennent en charge un service de surveillance qui permet aux administrateurs d'interroger un serveur NTP donné sur le nombre de hôtes connectés. Cette commande, appelée monlist, envoie au demandeur une liste des 600 derniers hôtes connectés au serveur interrogé. Comme décrit dans [84], un attaquant envoie à plusieurs reprises la demande « getmonlist » à un serveur NTP, tout en usurpant l'adresse IP du serveur demandeur par celle du serveur victime. Dans notre cas, nous pensons que les adresses de la source Internet, cherchent à identifier les serveurs ntp en prévision d'attaques futures, car le nombre de requête « getmonlist » envoyé pour chaque adresse source et 1(figure 34). (Voir l'annexe pour toutes les adresses sources) .

ET DOS Possible NTP DDoS Inbound Frequent Un-Authed MON_LIST Requests IMPL 0x03,"27.102.118.222","**.****.****.***",1

Figure 34.Message d'alerte NTP DDos

3.4 Scaning de réseau :

Les résultats de Bro révèlent 26 messages d'alertes concernant le Scaning du réseau, le message d'alerte indique clairement que la règle utilisée est l'analyse du même port pour plus de 25 hôtes. (Voir l'annexe pour tous les messages d'alertes).

222.186.129.44 scanned at least 25 unique hosts on port 22/tcp in 0m0s,remote,116

Figure 35.Message d'alerte Bro

3.5 Discussion :

Bien que notre analyse et les études précédentes confirment que la majorité du trafic darknet était du Scaning réseau, Bro ne mentionnait que 26 adresses sources. De plus, les résultats du Snort étaient assez différents de ceux de Bro.

Les pirates expérimentent et trouvent des moyens pour éviter le NIDS. En évitant NIDS, l'attaquant peut glisser sous le radar pour sonder des systèmes pouvant être surveillés par un NIDS sans détection, les plus importants sont :

- Analyses lentes : en analysant un réseau lentement, disons un port par heure, l'attaquant peut échapper à l'IDS en s'assurant que l'analyse dépasse la mémoire de l'IDS, de sorte que celle-ci n'est pas reconnue.
- Un changement de modèle : le NIDS s'appuie généralement sur la correspondance de modèle pour détecter une attaque. En modifiant légèrement les données utilisées dans l'attaque, il peut être possible d'éviter la détection.

4 Analyse temporelle :

Nous avons observé que le nombre de paquets TCP augmente sur la figure 26(section 2.1.1), en particulier pendant la période allant du 18 avril à 07h00 au 20 avril à 00h00, pour cela nous avons inspecté l'évolution du trafic par les 10 ports TCP les plus ciblés pour chaque journée (figure 36). Il apparaît clairement à la figure 37 que l'augmentation du trafic TCP est causée par l'augmentation de paquets TCP ciblant les ports 5061,5160,5038.

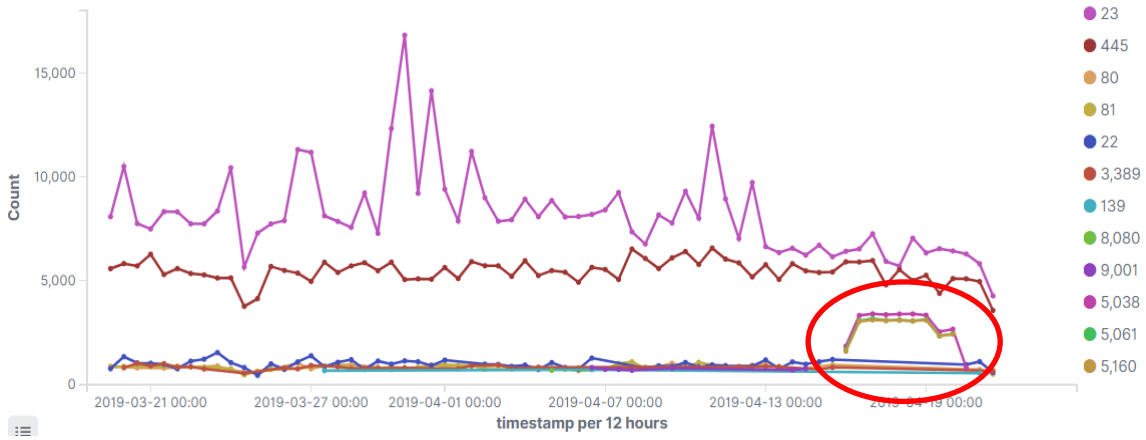


Figure 36.L'évolution de trafic par port

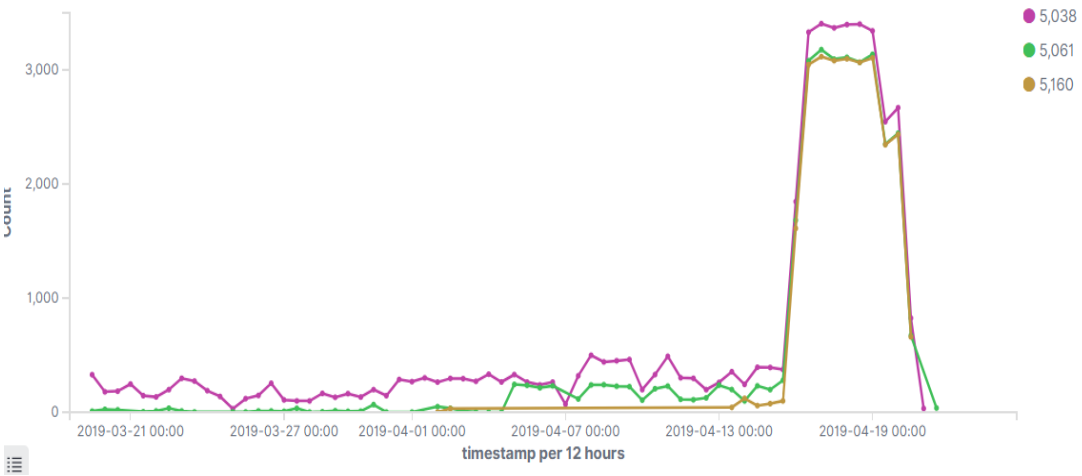


Figure 37.L'évolution de trafic pour les Ports (5038,5061,5160)

La figure 38 montre clairement que la source de ces paquets est l'adresses IP 87.251.81.86 (La Russie), qui n'a fait aucune activité sauf pendant cette période et n'a ciblé que les trois ports mentionnés.

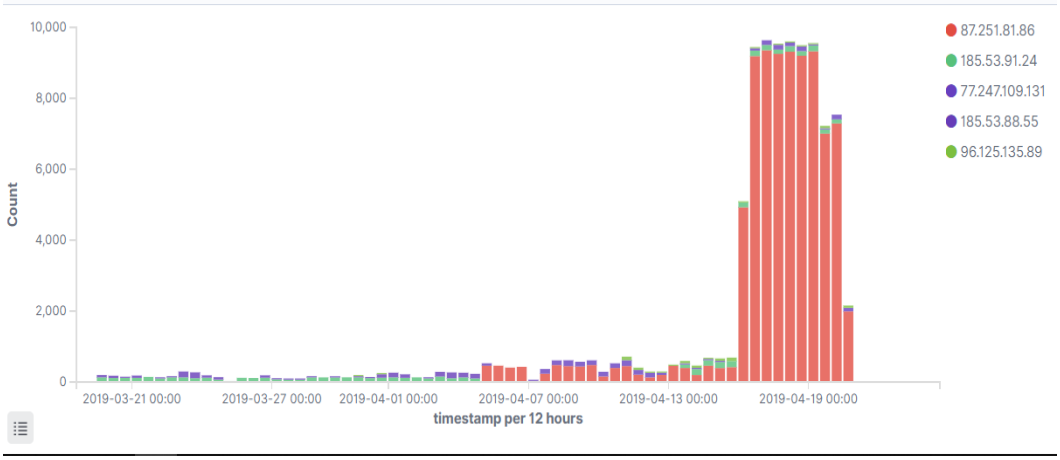


Figure 38.L'évolution du trafic par IP (Port 5061,5160,5038)

4.1 SIP Session Initiation Protocol :

Le protocole SIP (Session Initiation Protocol) est un protocole de contrôle de couche d'application qui peut établir, modifier et mettre fin aux sessions multimédia (conférences) telles que les appels téléphoniques Internet [83], également appelé voix sur IP (VoIP).

Une vulnérabilité non spécifiée dans les périphériques Cisco IOS 12.1 à 12.4 et 15.0 à 15.1, permet à des attaquants distants de provoquer un déni de service en envoyant des paquets SIP spécialement construits au ports TCP 5160,5061 [CVE-2011-3276], Un problème a été découvert sur les périphériques D-Link. Un attaquant peut exécuter du code arbitraire en injectant la commande shell dans le paramètre Sip de la page chkisg.htm. Cela permet un contrôle total sur les internes de l'appareil [CVE-2018-10823],

4.2 Nouvelle menace possible sur le port 5038

Le port 5038 non attribué par l'IANA, mais Asterisk Manager service utilise ce port 5038(Asterisk server est un autocommutateur téléphonique open source utilisée par les fournisseurs SIP VoIP), Il n'y a pas de vulnérabilité connue à ce jour sur ce port, Cependant, notre analyse indique qu'il existe une faille de sécurité que l'attaquant tente d'exploiter sur ce port et qui est toujours liée au protocole SIP.

5 Analyse approfondie :

Comme nous l'avons mentionné précédemment au chapitre 4, nous essayons de classer les adresses IP sources en plusieurs catégories en fonction du comportement des sources de trafic (IP source), afin de révéler les caractéristiques des comportements suspects, et des programmes malveillants inconnus.

Pour cela nous avons utilisé l'application que nous avons programmée, pour effectuer un clustering par l'algorithme K-means en utilisant la distance euclidienne comme mesure de similarité.

En raison de limitations hardware, nous avons analysé 538758 paquets collectés pendant la période au cours de laquelle il y avait une augmentation du trafic, pour ce trafic on prend on considération les premiers 1050 adresses IP unique dans le clustering, les clusters résultants contiendront tous les paquets provenant de ces adresses non seulement durant cette dernière période, mais également pendant toute la période de la collecte (voir le chapitre précédent).

5.1 Condition d'arrêt :

L'algorithme k-means est exécuté de manière itérative. Le critère d'arrêt indique à notre algorithme quand arrêter de mettre à jour les clusters. Pour assurer le meilleur clustering, nous arrêtons quand les centroïdes restent les mêmes après une itération, et pour assurer que l'itération donne le meilleur clustering, nous surveillons l'indice de Bouldin, pour chaque itération à travers l'application.

5.2 Choisir le nombre de clusters et les centroïdes :

Pour choisir le meilleur nombre de cluster, nous avons effectué le clustering des données, en partant de K=5 à 20 en comparant l'indice de Bouldin de chaque K, Le résultat est montré à la Figure 39.

Le meilleur K entre ces valeurs est 12 (La plus petite valeur de l'indice de Bouldin).

Pour éviter d'avoir des ensembles vides et d'essayer d'obtenir un clustering approprié les centroïdes sont sélectionnés au hasard, mais ils doivent être un vecteur de la matrice adresseIP/Port.

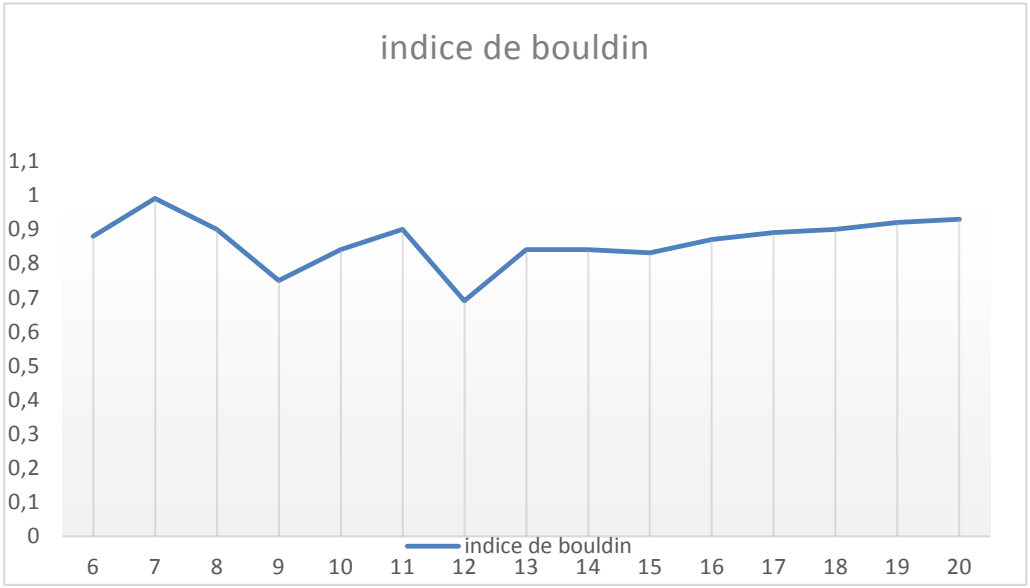


Figure 39.Indice de Bouldin (k=5 à 20)

5.3 Stabilité des clusters :

La stabilité du clustering est autre critère important. Si nous faisons le clustering plusieurs fois dans les mêmes conditions, est ce que nous obtenons les mêmes résultats . Dans notre cas, la réponse est oui, nous avons effectué le clustering pour k = 12 plusieurs fois, et le résultat était le même avec la même valeur d'indice de Bouldin.

5.4 Le résultat de clustering :

La figure 40 confirme que notre sélection de la condition d'arrêt était appropriée car la valeur la plus basse de l'indice de Bouldin était dans la dernière itération (9éme itération).

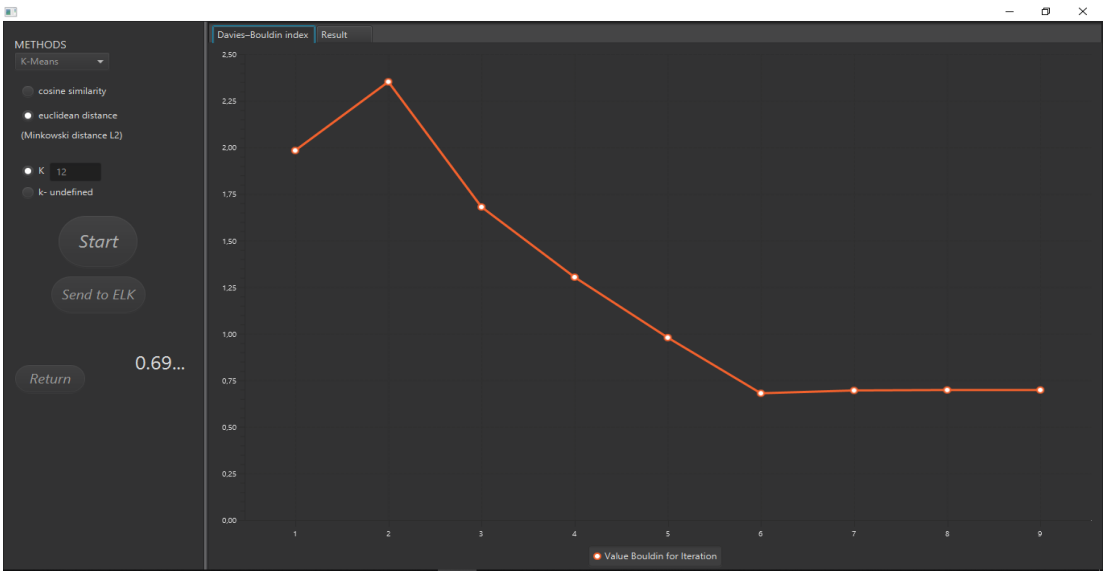


Figure 40.L'indice de Bouldin pour chaque itération (K=12)

Le résultat de clustering est montré dans le tableau 12.

Tableau 12.Le résultat de clustering

Cluster	Addresses IP	Paquets	Ports ciblés	conclusion
1	193.194.103.7 193.194.97.42	51740	7547; 23; 8291	Mirai Botnet
2	73 addresses	1099	445; 3389	
3	46.232.112.17	8669	(268) ports Ne font pas partie des 20 ports le plus ciblés	Il cherche de nouvelle vulnérabilité
4	87.251.81.86	88613	5060 ; 5160 ; 5038	SIP Voir la section 4
5	139.162.90.220	449	1723	Point-to-Point Tunneling Protocol vulnérabilité
6	139.162.71.210	328	8090	Utilisé comme alternative au port 8080, plusieurs trojans utilise ce port
7	125.64.94.212	2145	(52) Ports	//
8	172.105.226.61	378	9090	Plusieurs vulnérabilités
9	946 addresses	128066	(4955) Ports	Large scanning
10	193.32.161.48	2010	(77) Ports	//
11	18 addresses	1803	445	
12	202.29.57.103 107.170.105.134 106.13.106.176 129.204.105.36	14073	8545	JSON RPC default port, Il n'y a pas de vulnérabilité connue à ce jour sur ce port
Total	1050 addresses	299373	(5364) Ports	

5.4.1 Mirai : un botnet IoT

En septembre 2016, un nouveau ver a commencé à cibler les périphériques IoT embarqués sous Linux, tels que les caméras de vidéosurveillance IP. Telnet ou SSH sont souvent configurés sur ces périphériques avec les informations d'identification par défaut, mais ils n'utilisent pas activement ces services. Le malware Mirai utilise ces informations pour infecter les appareils. Les appareils infectés attendent des instructions de son serveur Command & Control pour lancer des attaques DDoS, Ce malware est responsable de la plus grande attaque de type DDoS à ce jour a eu lieu en février 2018. Cette attaque visait GitHub, un service de gestion de code en ligne populaire utilisé par des millions de développeurs, le trafic entrant pour cette attaque à un débit de 1,3 téraoctets par seconde (Tbps), envoyant des paquets à un débit de 126,9 millions par seconde [85]. Le code source de Mirai a divulgué après cette attaque de GitHub, ce code source est une référence pour définir les ports cibles de Mirai (23, 2323,5358) [86], bien qu'il existe de nombreuses variantes de Mirai disponibles ciblant d'autres ports (7547,8291).

5.5 Discussion des résultats :

Il est clair que la méthode que nous avons proposée a permis de séparer les adresses IP à comportement inhabituel (activités Scanning courantes), à l'exception du cluster 9, qui comprend la plupart des adresses IP, que nous classons comme activités de sondage larges et courantes, les autres clusters représentent des activités des adresses IP susceptibles d'être causées par des programmes malveillants connus (cluster 1 Mirai) ou inconnus. De plus, notre méthode classe l'adresse IP responsable de l'augmentation du trafic (voir section 4) seule dans le cluster 4, ce qui confirme l'efficacité de la méthode pour déterminer les comportements inhabituels.

5.6 Justification d'utilisation la Pondération PF-IPF :

Dans le chapitre 4, nous avons utilisé cette méthode pour à mettre en évidence un comportement de scanning inhabituel, pour justifier notre choix et démontrer son efficacité nous avons effectué le clustering sans l'utilisation de cette méthode, la meilleure valeur d'indice de Bouldin est 1.23 pour k=15. De plus les clusters résultants ne sont pas homogènes et ne représentent pas des activités inhabituelles .Cela confirme l'efficacité de cette méthode.

6 Conclusion :

Dans ce chapitre, nous avons étudié les données darknet en effectuant la caractérisation et le profilage du trafic. Nous avons interprété le résultat de cette étape en fournissant des informations sur les activités cybermenace. TCP représente le protocole le plus observé dans darknet. Le port TCP 23, relatif au service Telnet, est le port le plus ciblé. Le trafic de scanning constitue la majorité du trafic de réseau Darknet. Nous avons mis en évidence les menaces les plus pertinentes pour les ports les plus ciblés. Nous avons analysé le trafic avec NIDS. Cette étape d'analyse a révélé des menaces de gravité élevée, DDoS attaques, mais compte tenu de la quantité de trafic, leurs résultats sont jugés insatisfaisants. Pour cela, nous avons essayé d'extraire plus d'informations sur les menaces en appliquant des techniques de clustering. Les résultats ont montré l'efficacité de l'approche proposée, qui a permis de mettre en évidence des menaces connues (malware Mirai) et D'autres inconnues (des activités ciblant des ports avec aucun vulnérabilité connue à ce jour).

Conclusion générale :

Le cyberespace a favorisé l'émergence de nouvelles menaces et d'actes criminels. Pour garantir une meilleure sécurité de nos réseaux et systèmes il faut comprendre les motivations et les nouvelles techniques déployées par les attaquantes. Pour cela, il est efficace d'observer le trafic réseau et analyser les tendances d'attaques. Cela permet la détection précoce des menaces et la minimisation des dégâts.

Le Darknet composé d'un ensemble d'adresses IP non utilisées est devenu un outil populaire ces dernières années utilisé par les experts de la sécurité pour la surveillance du trafic réseau.

Dans cette optique notre objectif étant l'étude puis la mise en place d'un système de monitoring à base de Darknet pour la cybersécurité permettant la collecte et l'analyse des données pour la détection précoce des attaques.

Dans le présent travail nous avons mis en place le premier telescope réseau en Algérie au niveau du CERIST. Nous avons également mis en œuvre un système de monitoring basé sur ce telescope.

En raison de la grande quantité de données collectées à long terme, la conception du système de monitoring implique une analyse du processus de traitement et de stockage de ces données dans le but de tirer des conclusions fiables par rapport aux menaces potentielles. Pour cela, nous avons utilisé Wireshark pour capturer le trafic et le stack ELK pour le stockage, l'analyse et la visualisation des résultats.

L'analyse préliminaire basé sur les IDS et le profilage de données à l'aide de stack ELK nous a permis de caractériser la nature de trafic collectés et d'identifier certains comportements anormaux dans le trafic réseau en utilisant une analyse temporelle.

Afin de trouver une meilleure interprétation des données collectées nous avons procédé à une analyse approfondie dans laquelle nous avons proposé une approche basée sur les techniques de clustering k-means et euclidean distance.

L'application est écrite en Java, nous avons utilisé la bibliothèque JavaFx pour l'interface graphique, et pour faciliter l'interaction avec la base de données Elasticsearch nous avons utilisé la technologie JPA (Java Persistence API).

Les résultats obtenus sont très prometteurs et démontrent l'efficacité de l'approche proposée. Nous avons également implémenté les algorithmes k-median et, k-mediods et une classe pour les 2 mesures de similarités, cosinesimilarity et euclidean distance, cela nous donne 6 combinaisons possibles pour un futur travail.

Le télescope réseau ou Darknet du CERIST est important sur le plan académique. Malgré que les autres projets similaires dans le monde sont beaucoup plus grands car ils rassemblent diverses organisations telles que des universités, des fournisseurs d'accès Internet, des entreprises et même des ministères de la défense. Au niveau national, il n'existe aucune connaissance d'un projet similaire, ce qui en fait une référence importante dans ce contexte.

Enfin, cette première version du système Darknet au sein de CERIST répond aux objectifs escomptés, mais il est important de mentionner que pour optimiser son utilité, il est nécessaire d'avoir un processus d'actualisation et d'amélioration constante. Les versions suivantes et les nouvelles fonctionnalités mises en œuvre permettront de consolider progressivement le télescope pour en faire un système efficace de détection des menaces à la sécurité informatique.

Références

[1] V. Yegneswaran, P. Barford, and S. Jha. Global intrusion detection in the domino overlay system. In Proc. Network and Distributed System Security Symp. (NDSS), 2004

[2] Kim Y., Kim I., Park N. (2014) Analysis of Cyber Attacks and Security Intelligence. In: Park J., Adeli H., Park N., Woungang I. (eds) Mobile, Ubiquitous, and Intelligent Computing. Lecture Notes in Electrical Engineering, vol 274. Springer, Berlin, Heidelberg.

[3] Bou-Harb, Elias & Debbabi, Mourad & Assi, Chadi. (2014). Cyber Scanning : A Comprehensive Survey. Communications Surveys & Tutorials, IEEE. 16. 1496-1519. 10.1109 / SURV.2013.102913. 00020.

[4] Peng, Tao & Leckie, Christopher & Ramamohanarao, Kotagiri. (2007). Survey of network-based defense mechanisms countering the DoS and DDoS problems. ACM Comput. Surv.. 39. 10.1145/1216370.1216373.

[5] Rababah, Baha & Zhou, Shikun & Bader, Mansour. (2018). Evaluation the Performance of DMZ. I.J. Wireless and Microwave Technologies. 1. 1-13. 10.5815/ijwmt.2018.01.01.

[6]https://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xhtml [RFC 6335].

[7] Maglaras, Leandros & Ferrag, Mohamed Amine & Derhab, Abdelouahid & Mukherjee, Mithun & Janicke, Helge & Rallis, Stylianos. (2018). Threats, Countermeasures and Attribution of Cyber Attacks on Critical Infrastructures. Security and Safety. 5. 1-9. 10.4108/eai.15-10-2018.155856.

[8] David Salomon. (2010). Elements of computer security. Springer Science & Business Media.

[9] Abie, Habtamu. (2000). An Overview of Firewall Technologies.

[10] L. Hui and C. Yonghui. 2010 ."Research Intrusion Detection Techniques from the Perspective of Machine Learning," 2010 Second International Conference on Multimedia and Information Technology, Kaifeng, pp. 166-168.

[11] Othman, Suad & T Alsohybe, Nabeel & Mutaher Ba-Alwi, Fadl & Zahary, Ammar. (2018). Survey on Intrusion Detection System Types. 7. 444-462.

[12] Abdelkarim, Amjad & H. O. Nasereddin, Hebah. (2011). INTRUSION PREVENTION SYSTEM. INTERNATIONAL JOURNAL Of ACADEMIC RESEARCH. 3. 432-434.

[13] Wang, Zongjian & Li, Xiaobo. (2013). Intrusion Prevention System Design. 10.1007/978-1-4471-4847-0_47.

[14] Fachkha, Claude. (2015). Security Monitoring of the Cyber Space. 10.4018/978-1-4666-8456-0.ch004.

[15] Fachkha, Claude & Debbabi, Mourad. (2015). Darknet as a Source of Cyber Intelligence: Survey, Taxonomy and Characterization. IEEE Communications Surveys & Tutorials. 18. 1-1. 10.1109/COMST.2015.2497690.

[16] Mairh, Abhishek & Barik, Debabrat & Verma, Kanchan & Jena, Debasish. (2011). Honeypot in network security: A survey. ACM International Conference Proceeding Series. 600-605. 10.1145/1947940.1948065.

[17] Harrop, W & Armitage, G. (2005). Defining and Evaluating Greynets (Sparse Darknets). 344- 350. 10.1109/LCN.2005.46.

[18] Robert Petrunic, A.B.. (2015). Honeytokens as active defense. 1313-1317. 10.1109/MIPRO.2015.7160478.

[19] Michael Bailey, Evan Cooke, Farnam Jahanian, Andrew Myrick, and Sushant Sinha. Practical darknet measurement. In 40th Annual Conference on Information Sciences and Systems, pages 1496–1501. IEEE, 2006.

[20] Yu Jin, Zhi-Li Zhang, Kuai Xu, Feng Cao, and Sambit Sahu. Identifying and tracking suspicious activities through IP gray space analysis. In Proceedings of the 3rd annual ACM workshop on Mining network data, pages 7–12. ACM, 2007

[21] I. Polakis, G. Kontaxis, S. Ioannidis, and E. P. Markatos. Dynamic Monitoring of Dark IP Address Space (Poster). in International Work shop on Traffic Monitoring and Analysis. Springer, 2011, pp. 193–196.

[22] Michael Bailey, Evan Cooke, Farnam Jahanian, Jose Nazario, and David Watson. The Internet Motion Sensor: A distributed blackhole monitoring system. In Proceedings of Network and Distributed System Security Symposium (NDSS '05), San Diego, CA, February 2005.

[23] Zakir Durumeric, Michael Bailey, and J Alex Halderman. An internetwide view of internet-wide scanning. In USENIX Security Symposium,

2014.

[24] Rossow, Christian. (2014). Amplification Hell: Revisiting Network Protocols for DDoS Abuse. 10.14722/ndss.2014.23233.

[25] Fachkha, Claude & Bou-Harb, Elias & Debbabi, Mourad. (2014). Fingerprinting Internet DNS Amplification DDoS Activities. 10.1109/NTMS.2014.6814019.

[26] Xu, Ruomeng & Cheng, Jieren & Wang, Fengkai & Tang, Xiangyan & Xu, Jinying. (2019). A DRDoS Detection and Defense Method Based on Deep Forest in the Big Data Environment. Symmetry. 11. 78. 10.3390/sym11010078.

[27] Irwin, Barry. (2011). A framework for the application of network telescope sensors in a global IP network.

[28] Wustrow, Eric & Karir, Manish & Bailey, Michael & Jahanian, Farnam & Huston, Geoff. (2010). Internet Background Radiation Revisited. Proceedings of the ACM SIGCOMM Internet Measurement Conference, IMC. 62-74. 10.1145/1879141.1879149.

[30] Ruoming Pang, Vinod Yegneswaran, Paul Barford, Vern Paxson, and Larry Peterson (2004). Characteristics of Internet background radiation. In Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, pages 27–40.

[31] Irwin, Barry. (2013). A baseline study of potentially malicious activity across five network telescopes. 1-17.

[32] Fukuda, Kensuke & Hirotsu, Toshio & Akashi, Osamu & Sugawara, Toshiharu. (2009). Correlation Among Piecewise Unwanted Traffic Time Series. 1 - 5. 10.1109/GLOCOM.2008.ECP.314.

[33] Eduard Glatz and Xenofontas Dimitropoulos(2012). Classifying Internet one-way traffic. In Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE, pages 417–418.

[34] Wang, Ruoyu & Zhang, Ling & Zhen, Liu. (2013). A Novel Method of Filtering Internet Background Radiation Traffic. Proceedings - 4th International Conference on Emerging Intelligent Data and Web Technologies, EIDWT 2013. 371-376. 10.1109/EIDWT.2013.70.

[35] Cowie, Bradley & Irwin, Barry. (2010). Data classification for artificial intelligence construct training to aid in network incident identification using network telescope data. 356-360. 10.1145/1899503.1899544.

[36] Uli Harder, Matt W Johnson, Jeremy T Bradley, and William J

Knottenbelt.(2006). Observing internet worm and virus attacks with a small

network telescope. Electronic Notes in Theoretical Computer Science,

pages 47–59.

[37] Guofei Gu, Zesheng Chen, P. Porras, and Wenke Lee. (2007). Misleading and defeating importance-scanning malware propagation. In Third International Conference on Security and Privacy in Communications Networks. SecureComm., pages 250–259.

[38] Elias Bou-Harb, Mourad Debbabi, and Chadi Assi. (2013). A statistical approach for fingerprinting probing activities. In Eighth International Conference on Availability, Reliability and Security (ARES), pages 21– 30.

[39] Tomasz Andrysiak, Łukasz Saganowski, and Michał Choras. (2013). DDoS Attacks Detection by Means of Greedy Algorithms. In Image Processing and Communications Challenges, pages 303–310.Springer.

[40] Rajeev Gupta, Krithi Ramamritham, and Mukesh Mohania. (2013). Ratio threshold queries over distributed data sources. Proceedings of the VLDB Endowment, 6(8):565–576.

[41] David Moore, Colleen Shannon, et al. (2002). Code-red: a case study on the spread and victims of an internet worm. In Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment, pages 273–284.

[42] David Moore, Vern Paxson, Stefan Savage, Colleen Shannon, Stuart Staniford, and Nicholas Weaver. (2003) The Spread of the Sapphire/Slammer Worm. Technical report, CAIDA, ICSI, Silicon Defense, UC Berkeley EECS and UC San Diego CSE.

[43] David Moore, Vern Paxson, Stefan Savage, Colleen Shannon, Stuart Staniford, and Nicholas Weaver. (2003). Inside the slammer worm. IEEE Security & Privacy, pages 33–39.

[44] David Dagon, Cliff Changchun Zou, and Wenke Lee. (2006) Modeling Botnet Propagation Using Time Zones. In NDSS, volume 6, pages 2–13.

[45] Claude Fachkha, Elias Bou-Harb, and Mourad Debbabi. (2015) Inferring Distributed Reflection Denial of Service Attacks from Darknet. Computer Communications.

[46] Claude Fachkha, Elias Bou-Harb, and Mourad Debbabi. (2014) Fingerprinting Internet DNS Amplification DDoS activities. In 6th International Conference on New Technologies, Mobility and Security (NTMS), pages 1–5. IEEE.

[47] Jerome Francois, Olivier Festor, et al. (2006) Tracking global wide configuration errors. In IEEE/IST Workshop on Monitoring, Attack Detection and Mitigation.

[48] Craig Labovitz, Abha Ahuja, and Michael Bailey. (2001) Shining light on dark address space. Technical Report TR-2001-01, Arbor Netwoks, Ann Arbor, Michigan, USA.

[49] CAIDA: The UCSD Network Telescope. http://www.caida.org/projects/ network tele CAIDA. Conficker/Conflicker/Downadup as seen from the UCSD

[50] Network Telescope. http://www.caida.org/research/security/ms08-067/ conficker.xmlscope.

[51] Arbor Networks. ATLAS https://www.netscout.com/product/atlas intelligence-feed-aif

[52] Team Cymru: The Darknet Project https://www.team cymru.com/darknet.html

[53] Masashi Eto, Daisuke Inoue, Jungsuk Song, Junji Nakazato, Kazuhiro Ohtaka, and Koji Nakao. NICTER: a Large-Scale Network Incident Analysis System: Case Studies for Understanding Threat Landscape. In proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, BADGERS, pages 37–45, New York, NY, USA, 2011.

[54] Spiros Antonatos, Kostas Anagnostakis, and Evangelos Markatos.(2007) Honey@home: a new approach to large-scale threat monitoring. In Proceedings of the ACM workshop on recurring malcode, pages 38– 45.

[55] NoAH project. http://www.fp6-noah.org.

[56] Daisuke Inoue, Mio Suzuki, Masashi Eto, Katsunari Yoshioka, and Koji Nakao.(2009) DAEDALUS: Novel Application of Large-Scale Darknet Monitoring for Practical Protection of Live Networks. In Recent Advances in Intrusion Detection, pages 381–382. Springer.

[57] Japan cert coordination center. http://www.jpcert.or.jp

[58] Barry Vivian William Irwin.(2011) A framework for the application of network telescope sensors in a global IP network. PhD thesis, Rhodes University.

[59] Do Quoc Le, Taeyoel Jeong, H. Eduardo Roman, and James Won-Ki Hong. (2011).Traffic dispersion graph based anomaly detection. In Proceedings of the Second Symposium on Information and Communication Technology, SoICT, pages 36–41, New York, NY, USA.

[60] Cliff Joslyn, Sutanay Choudhury, David Haglin, Bill Howe, Bill

Nickless, and Bryan Olsen. (2013) Massive scale cyber traffic analysis: a

driver for graph database research. In First International Workshop

on Graph Data Management Experiences and Systems, page 3.

[61] Krasser, Sven & Conti, Gregory & Grizzard, Julian & Gribschaw, Jeff & Owen, Henry. (2005). Real-time and forensic network data analyis using animated and coordinated visualization. Proceedings from the 6th Annual IEEE System, Man and Cybernetics Information Assurance Workshop, SMC 2005. 2005. 42 - 49. 10.1109/IAW.2005.1495932.

[62] Fukuda, Kensuke & Fontugne, Romain. (2010). Estimating Speed of Scanning Activities with a Hough Transform. IEEE International Conference on Communications. 1-5. 10.1109/ICC.2010.5502264.

[63] Harrop, Warren & J. Armitage, Grenville. (2006). Real-time collaborative network monitoring and control using 3D game engines for representation and interaction. 31-40. 10.1145/1179576.1179583.

[64] Harrop, Warren & J. Armitage, Grenville. (2006). Modifying first person shooter games to perform real time network monitoring and control tasks. 10. 10.1145/1230040.1230074.

[65] Chindipha, Stones & Irwin, Barry & Herbert, Alan. (2018). Effectiveness of Sampling a Small Sized Network Telescope in Internet Background Radiation Data Collection.

[66] http://dev.maxmind.com/geoip/geoip2/geolite2/

[67] https://nmap.org/book/man-port-scanning-techniques.html

[68] Ceron, Joao & Steding-Jessen, Klaus & Hoepers, Cristine & Granville, Lisandro & Margi, Cintia. (2019). Improving IoT Botnet Investigation Using an Adaptive Network Layer. Sensors. 19. 727. 10.3390/s19030727.

[69] Microsoft Security TechCenter. Microsoft security bulletin ms09-018 - critical. http://technet.microsoft.com/en-us/security/bulletin/MS09-018.

[70] Saumil Shah. Top ten web attacks.URL http://www.blackhat.com/presentations/bh-asia-02/bh-asia-02-shah.pdf.

[71] Learn Security Online. MS terminal service cracking. URL http://www.carnal0wnage.com/papers/lso ms terminal server cracking.pdf

[72] J. Reynolds. ASSIGNED NUMBERS. RFC 1340, July 1992. URL https://tools.ietf.org/html/rfc1340. [Cited on page 9-26.]

[73] CVE-2016-3345. URL https://nvd.nist.gov/vuln/detail/CVE-2016-3345

[74] Ryba, Fabrice & Orlinski, Matthew & Wählisch, Matthias & Rossow, Christian & Schmidt, Thomas. (2015). Amplification and DRDoS Attack Defense -- A Survey and New Perspectives.

[75] Dulaunoy, Alexandre & Wagener, Gérard & Mokaddem, Sami & Wagner, Cynthia. (2017). AN EXTENDED ANALYSIS OF AN IOT MALWARE FROM A BLACKHOLE NETWORK Paper type Keywords.

[76] BadTunnel URL https://securityintelligence.com/badtunnel-bad-news/

[77] Fachkha, Claude & Bou-Harb, Elias & Boukhtouta, Amine & Dinh, Son & Iqbal, Farkhund & Debbabi, Mourad. (2012). Investigating the dark cyberspace: Profiling, threat-

based analysis and correlation. 7th International Conference on Risks and Security of Internet and Systems, CRiSIS 2012. 1-8. 10.1109/CRISIS.2012.6378947.

[78] RFC 792 ICMP URL: https://tools.ietf.org/html/rfc792

[79] RFC 791 INTERNET PROTOCOL URL :https://tools.ietf.org/html/rfc791

[80] The Morgan Kaufmann Series in Data Management Systems: Data Mining. Concepts and Techniques, [3rd Edition] (2011).Chapitre 10.

[81] Bou-Harb, Elias & Debbabi, Mourad & Assi, Chadi. (2014). On Fingerprinting Probing Activities. Computers & Security. 43. 10.1016/j.cose.2014.02.005.

[82] Nishikaze, Hironori & Ozawa, Seiichi & Kitazono, Jun & Ban, Tao & Nakazato, Junji & Shimamura, Jumpei. (2015). Large-Scale Monitoring for Cyber Attacks by Using Cluster Information on Darknet Traffic Features. Procedia Computer Science. 53. 175-182. 10.1016/j.procs.2015.07.292

[83]Bou-Harb, Elias & Lakhdari, Nour-Eddine & Binsalleeh, Hamad & Debbabi, Mourad. (2014). Multidimensional investigation of source port 0 probing. Digital Investigation.

[84] CVE-2013-5211 URL:https://nvd.nist.gov/vuln/detail/CVE-2013-5211

[85] Ben, Najah & Biondi, Fabrizio & Bontchev, Vesselin & Decourbe, Olivier & Given-Wilson, Thomas & Legay, Axel & Quilbeuf, Jean. (2018). Detection of Mirai by Syntactic and Behavioral Analysis. 224-235. 10.1109/ISSRE.2018.00032.

[86] Dulaunoy, Alexandre & Wagener, GÃ©rard & Mokaddem, Sami & Wagner, Cynthia. (2017). AN EXTENDED ANALYSIS OF AN IOT MALWARE FROM A BLACKHOLE NETWORK Paper type Keywords.

Annexes

PROTOCOLE :

Les trois protocoles les plus courants observés dans cette étude ont été présentés dans le tableau

Numéro de protocole	Keyword	Nom	Références
1	ICMP	Internet Control Message	RFC792
6	TCP	Transmission Control	RFC793
17	UDP	User Datagram	RFC768

Fichier configuration logstash :

```
input {
  file {
    path => "C:/Users/yacine/Documents/M2/S2/Trafic/cap/dark7.json"
    start_position => "beginning"
  }
}
filter {
    #Drop Elasticsearch Bulk API control lines
    if ([message] =~ "{\"index") {
        drop {}
      }
    json {
```

```
        source => "message"
        remove_field => "message"
    }
    #Extract innermost network protocol
    grok {
        match => {
            "[layers][frame][frame_frame_protocols]" => "%{WORD:protocol}$"
        }
    }

    date {
        match => [ "timestamp", "UNIX_MS" ]
    }
  geoip {

        source => "layers.ip.ip_ip_src"
        target => "geoip"

      }

}

output {
```

```
elasticsearch {
hosts => "localhost:9200"
index => "packets-%{+YYYY-MM-dd}"
document_type => "_doc"
manage_template => false
pipeline => "geoip"

  }
#stdout {
 #     codec =>rubydebug}
}
```

TEMPLATE Packets ELASTICSEARCH:

```
PUT _template/packets
{
  "template": "packets-*",
  "mappings": {
    "pcap_file": {
      "dynamic": "false",
      "properties": {
        "timestamp": {
          "type": "date"
```

```
    },
    "layers": {
      "properties": {
        "icmp": {
          "properties": {
            "icmp_type": {
              "type": "integer"
            },
            "icmp_code": {
              "type": "integer"
            }
          }
        },
        "ip": {
          "properties": {
            "ip_ip_src": {
              "type": "ip"
            },
            "ip_ip_dst": {
              "type": "ip"
            },
            "ip_id": {          //identifie de manière unique le fragment d'un paquet IP
d'origine
```

```
            "type": "integer"
          },
          "ip_ttl": {    // time to live
            "type": "integer"
          },
          "ip_proto": {   //
            "type": "integer"
          },
          "ip_tos": {   // type of service ou dscp maintenant , qui peut être utilisée pour
empêcher
                                l'abandon de paquets pendant les périodes d'encombrement
du réseau.
            "type": "integer"
          },
          "ip_len": {      // Longueur totale  y compris l'en-tête et les données
            "type": "integer"
          },
          "ip_flags_df": {    //do not fragement, Le bit DF =1  pour désactiver la
fragmentation
            "type": "boolean"
        },
        "ip_flags_mf": {   //plus de fragement (more fragemen ) le dernier fragement =0
          "type": "boolean"
        },
```

```
    "ip_frag_offset": { //indique la position dans laquelle placer le fragment de paquet
                                                pour  reconstituer le paquet d'origine
      "type": "integer"
    }
   }
},
"udp": {
  "properties": {
    "udp_udp_srcport": {
      "type": "integer"
    },
    "udp_udp_dstport": {
      "type": "integer"
    }
  }
},
"tcp": {
  "properties": {
    "tcp_tcp_srcport": {
      "type": "integer"
    },
    "tcp_tcp_dstport": {
      "type": "integer"
```

},

"tcp_flags_tcp_flags_urg": {

"type": "integer"

},

"tcp_flags_tcp_flags_ack": {

"type": "integer"

},

"tcp_flags_tcp_flags_push": {

"type": "integer"

},

"tcp_flags_tcp_flags_reset": {

"type": "integer"

},

"tcp_flags_tcp_flags_syn": {

"type": "integer"

},

"tcp_flags_tcp_flags_fin": {

"type": "integer"

},

"tcp_tcp_seq": {

"type": "integer"

},

"tcp_tcp_ack": {

```
            "type": "integer"
          },
          "tcp_tcp_window_size": {
            "type": "integer"
          }
        }
      }
    },
    "geoip": {
      "dynamic": "true",
      "properties": {
        "location": {
          "type": "geo_point"
        }
      }
    }
  }
 }
}
```

PIPELINE GEOIP :

```
PUT  _ingest/pipeline/geoip
```

```
{
  "description" : "Add GeoIP Info",
  "processors" : [
    {
      "geoip" : {
        "field" : "layers.ip.ip_ip_src"
      }
    }
  ]
}
```

Les messages d’alertes Bro :

Message,"Sub-Message",Count

222.186.129.44 scanned at least 25 unique hosts on port 22/tcp in 0m0s,remote,116

115.238.245.8 scanned at least 25 unique hosts on port 22/tcp in 0m0s,remote,101

110.249.212.46 scanned at least 25 unique hosts on port 3128/tcp in 0m0s,remote,83

221.194.44.151 scanned at least 25 unique hosts on port 1433/tcp in 0m0s,remote,55

103.237.145.146 scanned at least 25 unique hosts on port 9090/tcp in 0m0s,remote,42

222.186.160.100 scanned at least 25 unique hosts on port 60001/tcp in 0m0s,remote,36

185.232.67.11 scanned at least 25 unique hosts on port 22/tcp in 0m0s,remote,28

110.249.212.46 scanned at least 25 unique hosts on port 8118/tcp in 0m0s,remote,17

58.218.66.177 scanned at least 25 unique hosts on port 60001/tcp in 0m0s,remote,16

187.115.165.204 scanned at least 25 unique hosts on port 22/tcp in 0m0s,remote,13

101.230.200.173 scanned at least 25 unique hosts on port 5900/tcp in 0m0s,remote,12

58.218.213.79 scanned at least 25 unique hosts on port 3306/tcp in 0m0s,remote,12

103.85.84.85 scanned at least 25 unique hosts on port 60001/tcp in 0m0s,remote,11

94.26.234.12 scanned at least 25 unique hosts on port 4145/tcp in 0m0s,remote,11

123.129.217.29 scanned at least 25 unique hosts on port 3306/tcp in 0m0s,remote,10

185.232.67.101 scanned at least 25 unique hosts on port 22/tcp in 0m0s,remote,10

206.189.181.86 scanned at least 25 unique hosts on port 5900/tcp in 0m0s,remote,10

62.149.99.199 scanned at least 25 unique hosts on port 445/tcp in 0m0s,remote,10

113.4.133.5 scanned at least 25 unique hosts on port 1433/tcp in 0m0s,remote,9

113.4.133.5 scanned at least 25 unique hosts on port 2433/tcp in 0m0s,remote,9

58.218.213.79 scanned at least 25 unique hosts on port 60001/tcp in 0m0s,remote,9

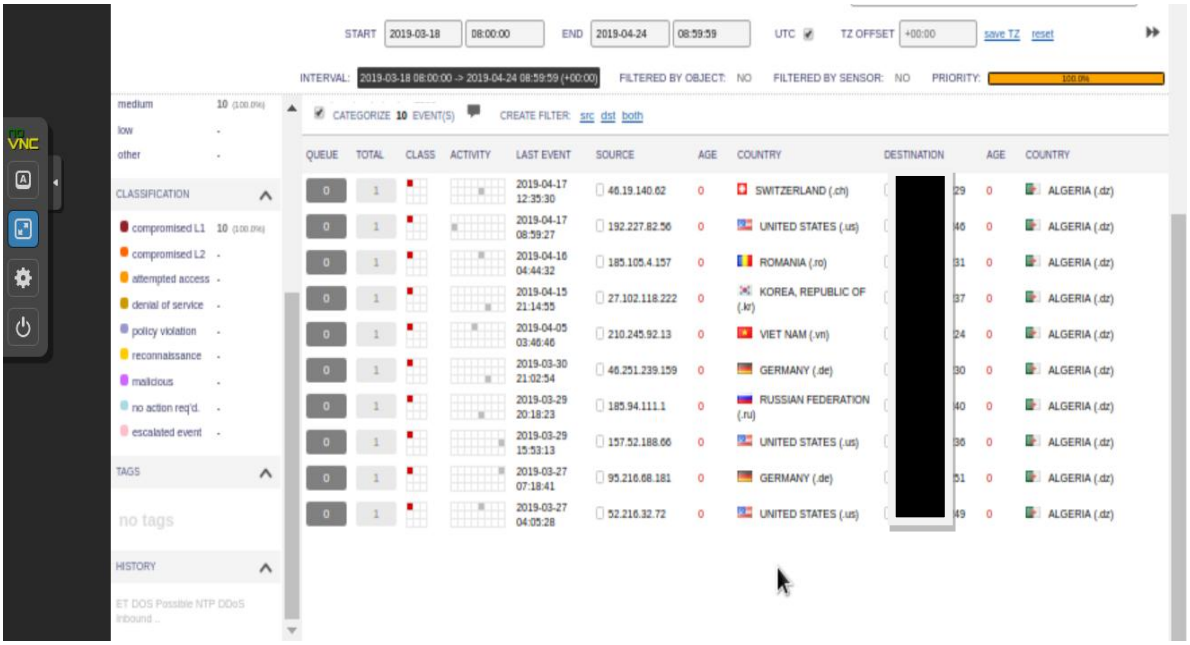
101.226.175.133 scanned at least 25 unique hosts on port 445/tcp in 0m0s,remote,8

113.4.133.5 scanned at least 25 unique hosts on port 1444/tcp in 0m0s,remote,8

139.220.192.57 scanned at least 25 unique hosts on port 22/tcp in 0m0s,remote,8

45.254.26.13 scanned at least 25 unique hosts on port 60001/tcp in 0m0s,remote,8

Snort:


START 2019-03-18 08:00:00 END 2019-04-24 08:59:59 UTC TZ OFFSET +00:00 save TZ reset
INTERVAL: 2019-03-18 08:00:00 -> 2019-04-24 08:59:59 (+00:00) FILTERED BY OBJECT: NO FILTERED BY SENSOR: NO PRIORITY:
medium 10 (100.0%)
low -
other -
CLASSIFICATION
compromised L1 10 (100.0%)
compromised L2 -
attempted access -
denial of service -
policy violation -
reconnaissance -
malicious -
no action req'd. -
escalated event -
TAGS
no tags
HISTORY
ET DOS Possible NTP DDoS Inbound ...
CATEGORIZE 10 EVENT(S) CREATE FILTER: src dst both
QUEUE TOTAL CLASS ACTIVITY LAST EVENT SOURCE AGE COUNTRY DESTINATION AGE COUNTRY
0 1 2019-04-17 12:35:30 46.19.140.62 0 SWITZERLAND (.ch) 29 0 ALGERIA (.dz)
0 1 2019-04-17 08:59:27 192.227.82.56 0 UNITED STATES (.us) 46 0 ALGERIA (.dz)
0 1 2019-04-16 04:44:32 185.105.4.157 0 ROMANIA (.ro) 31 0 ALGERIA (.dz)
0 1 2019-04-15 21:14:55 27.102.118.222 0 KOREA, REPUBLIC OF (.kr) 37 0 ALGERIA (.dz)
0 1 2019-04-05 03:46:46 210.245.92.13 0 VIET NAM (.vn) 24 0 ALGERIA (.dz)
0 1 2019-03-30 21:02:54 46.251.239.159 0 GERMANY (.de) 30 0 ALGERIA (.dz)
0 1 2019-03-29 20:18:23 185.94.111.1 0 RUSSIAN FEDERATION (.ru) 40 0 ALGERIA (.dz)
0 1 2019-03-29 15:53:13 157.52.188.66 0 UNITED STATES (.us) 36 0 ALGERIA (.dz)
0 1 2019-03-27 07:18:41 95.216.68.181 0 GERMANY (.de) 51 0 ALGERIA (.dz)
0 1 2019-03-27 04:05:28 52.216.32.72 0 UNITED STATES (.us) 49 0 ALGERIA (.dz)