

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Centre De Recherche Sur L'information Scientifique Et Technique



Mémoire

Présenté en vue de l'obtention du diplôme de poste graduation spécialisé
Informatique

Option : Big Data et calcule intensif

Thème :

**Le Monitoring Des Cyber-Attaques Basé Sur Le Darknet
En Utilisant Les Technologies Big Data**

Réalisé par :

NAFAA Mohammed Sabri

RHASKALI Abdennacer

Soutenu devant le jury composé de :

Président : - Mr Belazzougui Djamel CERIST

Examinatrice : - Mme Guemraoui Lila CERIST

Encadreur : - Mme NOUALI-TABOUDJEMAT Nadia CERIST

- Melle ZEGHACHE Linda CERIST.

2018-2019

Remerciements

Tout d'abord nous remercions le bon Dieu de nous avoir donné la force et la volonté de débiter et de terminer ce modeste travail, tout en espérant qu'il soit bon et acceptable.

Mes remerciements les plus distingués vont, en premier lieu, à mon encadreur

Dr.NOUALI-TABOUDJEMAT Nadia et Co-encadreur

Dr.ZEGHACHE Linda pour leur disponibilité et leur encouragement tout au long de la réalisation de ce mémoire.

Mes remerciements s'adressent également aux membres du jury qui m'ont fait l'honneur d'accepter d'évaluer ce travail.

Je remercie également toute personne ayant contribué de près ou de loin à l'élaboration de ce travail de recherche. Je tiens à remercier le personnel pédagogique et administratif du CERIST, pour leur aide précieuse et leur large disponibilité.

Dédicace

Tous ceux et toutes celles qui nous ont soutenu tout au long de la préparation de ce travail et qui nous ont encouragé à le poursuivre et à l'achever, méritent dédicace de ce mémoire.

Je le dédie à :

Mes très chers parents : pour leur amour, soutien et énormes sacrifices.

A mon épouse pour son soutien et encouragement

A mes adorables fils Mohamed Nassim et Abd Errahmen

A mes chers frères et sœurs

A toute la famille et les amies qui m'ont moralement soutenu le long de ce travail de recherche.

Résumé

Traditionnellement, les technologies et les outils utilisés pour le contrecarrer les cyber-menaces ont été plus réactifs que proactifs. Les cyber-menaces sont devenues de plus en plus complexes et difficiles à contrer. Cela a conduit les entreprises à revoir et à renouveler les moyens et les pratiques de la cyber-sécurité classique en adaptant de nouvelles technologies pour passer d'un système réactif à un système de proactif afin d'améliorer la sécurité.

Dans cette optique, nous nous sommes intéressés aux télescopes réseaux communément appelé Darknets, qui sont des systèmes de monitoring à base de piège permettant d'extraire des renseignements sur les cyber-menaces. Une analyse complexe sur Les données Darknet permet de révéler des prémisses d'attaques et d'assurer en plus de la détection la prévention contre les cyber-menaces.

Afin d'analyser les données Darknet plus rapidement et en temps quasi réel nous avons intégré les outils Big Data. Les outils analytiques du Big Data offrent la capacité d'analyser différents types de données en provenance de sources diverses et de réagir en temps réel. Ces outils ne permettent pas seulement de rassembler des informations, mais aussi de connecter ces données, et d'établir des corrélations et des connexions. Il est ainsi possible d'augmenter l'efficience et de contrecarrer plus facilement les cyber-attaques.

Dans notre travail nous avons analysé des données collectées par le Darknet du CERIST sous forme de fichiers Pcap, et les avons intégrés sur dans une plateforme Big Data. Nous avons défini un pipeline composé de plusieurs étapes de traitements utilisant chacune un outil Big Data adéquat afin de préparer les données à l'analyse et la visualisation.

Enfin, nous avons effectués des analyses et généré des aperçus en utilisation Elasticsearch pour l'indexation et Kibana comme tableau de bord pour la visualisation.

Mots clés : Big Data, Spark, Darknet, ELK, cyber-Attaques.

ملخص

كانت التقنيات والأدوات التقليدية المستخدمة لإحباط التهديدات الإلكترونية أكثر تفاعلية من كونها استباقية، لكن التهديدات السيبرانية أصبحت معقدة بشكل متزايد ويصعب مواجهتها مما دفع بالشركات لمراجعة وتجديد وسائل وممارسات الأمن السيبراني القديمة من خلال تكييف تقنيات جديدة للانتقال إلى نظام استباقي.

و للغرض السالف الذكر، وضعنا كل الاهتمام على تلسكوبات الشبكة المعروفة باسم Darknets ، وهي أنظمة مراقبة تعتمد أساسا على فح لتمكننا من استخراج معلومات حول التهديدات السيبرانية. إن التحليل المعقد لبيانات Darknet يمكننا من معرفة مصدر الهجمات، بالإضافة إلى التأكد و الكشف وتفادي التهديدات الإلكترونية.

قصد تحليل بيانات Darknet بشكل أسرع وفي الوقت الفعلي تقريبا، قمنا بدمج أدوات البيانات الضخمة من أجل توفير القدرة على تحليل أنواع مختلفة من البيانات من مصادر مختلفة، والتفاعل في الوقت الفعلي. كما أنها لا تقوم بجمع المعلومات معًا فحسب، بل تعمل أيضًا على توصيل تلك البيانات وإنشاء الارتباطات والوصلات، مما يمكن زيادة الكفاءة ومنع الهجمات الإلكترونية بسهولة أكبر.

في عملنا هذا، قمنا بتحليل البيانات التي تم جمعها بواسطة Darknet الخاص ب CERIST في شكل ملفات Pcap ، ودمجناها في أرضية البيانات الضخمة، أين حددنا مخطط يتكون من عدة خطوات معالجة، كل منها يستخدم أداة مناسبة لإعداد البيانات لمرحلي التحليل والتصوير.

أخيرًا ، أجرينا التحليل وتوصلنا إلى رؤى باستخدام Elasticsearch للفهرسة و Kibana كلوحة تحكم للتصور.

الكلمات المفتاحية : Big Data ، Spark ، Darknet ، ELK ، cyber-Attacks.

Summary

The traditional technologies and tools used to thwart electronic threats have been more reactive than proactive, but cyber threats are becoming increasingly complex and difficult to confront, prompting companies to review and renew old cybersecurity tools and practices by adapting new technologies to transition to a proactive system.

For the aforementioned purpose, we have put all our attention on network telescopes known as Darknets, which are surveillance systems that rely mainly on a trap to enable us to extract information about cyber threats. Darknet's intricate analysis of the data enables us to know the source of the attacks, as well as to confirm, detect and avoid cyber threats.

Intended to analyze Darknet data faster and in near real time, we have integrated big data tools in order to provide the ability to analyze different types of data from different sources, and interact in real time. Not only does it bring information together, but it also delivers that data and creates connections and connections, which can increase efficiency and prevent cyber attacks more easily.

In our work, we analyzed the data collected by CERIST's Darknet in the form of Pcap files, and integrated them into the Big Data floor, where we identified a chart consisting of several processing steps, each of which uses a suitable tool to prepare the data for the analysis and visualization stages.

Finally, we ran the analysis and came up with insights using Elasticsearch for indexing and Kibana as a dashboard for visualization.

Keywords : Big Data, Spark, Darknet, ELK, cyber-Attacks.

Sommaire:

Remerciement

Résumé

Introduction générale.....01

Chapitre 1 : Les systèmes de monitoring à base de piège pour la cyber sécurité

I.	Introduction	03
II.	Les Systèmes de monitoring pour la cyber sécurité	03
	1. Pare-feu (Firewall)	03
	2. Système de détection d'intrusion (IDS)	04
	3. Système de prévention d'intrusion (IPS)	04
III.	Les Systèmes de monitoring à base de pièges	04
	1. Darknet	05
	2. IP Gray Space	06
	3. Honeypots (Les pots de miel)	06
	4. Greynet	07
	5. Honeytokens	07
	6. Comparaisons	08
IV.	Darknet source pour la cyber intelligence	09
	1. Les types de menaces détectées par le Darknet	10
	1.1 Activités de scan (Probing/Scanning)	10
	1.2 Les attaques des deni de services distribués (DDoS)	10
	1.3 Les attaques DRDoS	11

2. Données Darknet	12
2.1 BACKSCATTER	12
2.2 MISCONFIGURATION	12
2.3 AGRESSIF / HOSTILE	12
3. Techniques d'analyse de données.....	13
3.1 Profilage des données (Data Profiling)	13
4. Filtrage et classification des données	13
V. Conclusion.....	14

Chapitre 2 : Big Data ; outils et qualités

I. Introduction	15
II. Définition du Big Data	15
1. Littéralement.....	15
2. Conceptuellement.....	15
3. Généralement.....	15
III. les caractéristiques du Big Data	16
1. Volume	16
2. Vitesse	16
3. Variété	16
4. Véracité	17
5. Valeur	17
IV. Architecture Big Data	17
1. Intégration	17
2. Stockage de données (Data Storage)	17

3. traitement de données (Data Processing)	18
4. Sécurité	18
5. Opération	18
V. Les plateformes de Big Data	18
VI. Apache Hadoop.....	18
1. HDFS (Hadoop Distributed File System):partie stockage	19
2. MapReduce	20
2.1. Principe MapReduce	21
2.2. Les caractéristiques de MapReduce	21
3. YARN, partie management	22
4. Apache Spark	23
4.1-L'architecture de Spark.....	23
a/ Le stockage.....	23
b/ L'API	23
c/Gestion des ressources	23
VII. Conclusion	25

Chapitre 3 : Conception

I. Introduction	26
II. Architecture générale.....	26
III. Source de données	27
IV. La description de notre architecture.....	28
1-Chargement de données	28
2-Traitement de données	29

2.1 Les étapes de traitement	30
2.1.1 Le nettoyage.....	30
2.1.2 L'enrichissement	30
2.1.3 L'ingestion	30
3-Indexation et visualisation	31
3.1 -indexation	31
3.1.1 Un index Elasticsearch	32
3.1.2 Qualité d'Elasticsearch	33
3.1.3 Fonctionnement d'Elastic search	33
3.2 Visualisation	33
V. Conclusion	35

Chapitre 4 : Implémentation

I- Introduction.....	36
II- Schéma général du pipeline Big Data implémenté.....	36
III- Les ressources matérielles et logicielles utilisées.....	37
1- Ressources matérielles.....	37
2- Logiciels utilisés	38
IV- Préparations des données.....	38
1- Structure de données.....	38
2- Chargement de données.....	40
V- Les étapes de traitement	40
1- Le nettoyage.....	41
2- L'enrichissement.....	41
3- L'ingestion.....	42

VI-	Indexation, visualisation et analyse.....	43
1-	Indexation.....	43
2-	Visualisation.....	44
1-	Graphe du trafic.....	44
2-	Distribution des protocoles.....	45
3-	Les ports TCP les plus ciblés.....	45
4-	Les ports UDP les plus ciblés.....	46
5-	Le tableau des flags TCP	47
6-	La géolocalisation.....	48
VII-	Top 20 des adresses IP sources.....	49
VIII-	Conclusion.....	50
	Conclusion générale.....	51
	Références bibliographiques	53

Liste des figures

Figure 01 : concept de base de capteur de surveillance basé sur des pièges.....	05
Figure 02 : Activités de Probing.....	10
Figure 03 : Activités DDoS.....	11
Figure 04 : Activités DRDoS.....	12
Figure 05 : Les 5V qui caractérisent les Big Data.	16
Figure 06: Architecture Big Data.....	17
Figure 07: Affichage graphique de partition et de stockage des fichiers dans HDFS.....	19
Figure 08 : Présentation du processus d'écriture d'un fichier sur HDFS.....	20
Figure 09 : Fonctionnement de MapReduce.....	21
Figure 10 : Yarn.....	23
Figure11: Architecture Spark.....	24
Figure 12 : Architecture générale.....	27
Figure 13 : Chargement de données.....	28
Figure 14 : Traitement de données.....	39
Figure 15 : Transfert de données vers Hbase.....	31
Figure 16 : Processus d'indexation.....	32
Figure 17 : Processus de visualisation.....	34
Figure 18 : L'implémentation de notre flux Big Data.....	37
Figure 19 : Le nettoyage des données.....	41

Figure 20 : Création d'un indexe.....	43
Figure 21 : L'indexation	43
Figure 22 : volume de trafic par heure.....	44
Figure 23 : distribution des protocoles	45
Figure24: Top 10 destination Ports TCP	46
Figure25: Top 10 destination Ports UD.....	47
Figure 26: Localisation géographique en fonction du volume de trafic provenant de chaque pays.....	48
Figure 27: les top 20 adresses IP sources.....	49

Liste des tableaux

Tableau 1 : Systèmes de surveillance basés sur des pièges .Comparaison.....	08
Tableau 2 : Caractéristique du matériel utilisé.....	37
Tableau 3 : Caractéristique des Logiciels utilisés	38
Tableau 4 : Les combinaisons des indicateurs (TCP flags) utilisés dans les paquets TCP..	48