TABLE DES MATIERES

DEDICACES : i
REMERCIMENTS : ii
TABLE DES MATIERES iii
GLOSSAIRE: vi
AVANT PROPOS : viii
RESUME : ix
ABSTRACT: ix
INTRODUCTION GENERALE : 1
Chapitre 1 : Cadre théorique et méthodologique : 3
I. Cadre théorique 3
I-1-Définition du besoin 3
I-2-Importance de la question 3
I-3-Formulation des objectifs 4
II. Méthodologie 4
II-1-Délimitation du sujet : 4
II-2-Les techniques d'investigation : 4
II-3-Echantillonnage : 5
II-4-Les difficultés rencontrées : 6
Chapitre 2 : Sécurité des systèmes d'information 7
III. Généralités sur la sécurité 7
III-1-Les différents aspects de la sécurité 8
III-2-La politique de sécurité 8
IV. Différentes solutions de sécurité : 10
IV.1 Les Firewall (Pare-feu) : 10
a- Fonctionnement d'un pare-feu : 11
b- Le filtrage simple de paquets : 11
c- Le filtrage dynamique 12
d- Le filtrage applicatif : 13
e- Les limites des pare-feu : 13
IV.2 Les systèmes de détection et de prévention d'intrusion : 14
a- IDS : 14
b- Les IPS : 18

IV.3 La supervision du réseau 19
IV.4 La gestion des Logs 20
Chapitre 3 : L'utilisation des outils SIEM pour la sécurité d'un réseau : 22
V. Généralité sur les SIEMs 22
V.1 Définition d'un SIEM 22
V.2 Principe de fonctionnement des SIEM: 24
V.3 Les coûts : 26
V.4 Les prérequis à la mise en place d'un SIEM : 28
V.5 La sécurité de l'application SIEM : 31
V.6 Architecture de gestion des informations et des événements de sécurité (SIEM) : 33
V.7 Avantages et limites des SIEM : 35
V.8 Les tendances futures du SIEM : 37
V.9 Différence entre les SIEM et la gestion des logs : 38
VI. Solutions SIEM 39
VI-1-Comparaison entre les outils SIEM et la Gestion des logs 39
VI-2-Les principales solutions SIEM les plus utilisées du marché : 40
a-Micro Focus ArcSight : 40
b- Datadog : 40
c-AT&T Cyber Security (AlienVault anciennement OSSIM : Open Source Security Information Management) : 41
d-SolarWinds : 49
e-QRadar : 50
f-Elasticsearch – Logstash – Kibana (ELK) : 54
g-Logstash : 55
h-Splunk : 57
VI-3-Tableau comparatif des outils SIEM étudiés dans ce travail : 64
VI-4-Choix de la solution : 65
Chapitre 4 : Déploiement de la solution et tests 66
VII. Déploiement de Splunk Entreprise Security : 66
VII-1-Architecture de Splunk ES : 66
VII-2-Différentes étapes du pipeline de données : 66
VII-.3-Composants Splunk : 68
V II-8-Modes de déploiement et Configuration requise de Splunk ES 71
VII-9-Déploiement de Splunk ES 73

a-Préparation du déploiement : 74
b- Installations des services sur les serveurs : 74
c- Installation de Splunk Entreprise Security sur la machine Splunk-Indexer : 74
d-Installation de Splunk Universal Forwarder sur la machine WEB-SRVR 75
e-Déploiement de Splunk Entreprise Security : 76
f-Ingestion de données : 77
VII-10-La fonction Research de Splunk ES : 81
VIII. Tests sur quelques domaines d'applications de Splunk ES 90
VIII-1-Exemple de surveillance avec Splunk : 90
VIII-2-Exemple de recherche de logs sur le serveur Web : 91
CONCLUSION GEENERALE : 93
BIBLIOGRAPHIE: 94
WEBOGRAPHIE: 95

GLOSSAIRE:

APT : Advanced Persistent Treat

AT&T : American Telephone & Telegram

AWS : Amazone Web Service

BI : Business Inteligence

CID : Confidentialité Intégrité Disponibilité

CPU : Central Processing Unit

DISA : Defense Information System Agency

DMZ : Demilitarized Zone

DSS : Data Security Standard

E-streamer : System Event Streamer

GHz : Giga Hertz

Go : Giga octet

HIDS : Host Based IDS

HIPAA : Health Insurance Port ability and Accountability Act

HTTP : HyperText Transfert Protocol

IA : Intelligence Artificielle

ICMP : Internet Control Message Protocol

ID : IDentifier

IDMEF3 : Intrusion Detection Detection Message Exchange Format3

IDS : Intrusion Detection System

IP : Internet Protocole

LEA : (OPSEC) Log Export API

NIDS : Network-Based IDSs

OWASP : Open Web Application Security Project

PCI DSS : Payement Cards Industry Data Security Standard

PKI : Public Key Infrastructure

PME : Petite et Moyenne Entreprise

RAM : Random Access Memory

SDEE : Security Device Event Exchange

SIEM : Security Information and Event Management

AVANT PROPOS :

La sécurité des données est à la préoccupation d'un bon nombre d'entreprises. De ce fait il est fondamental de penser à la formation des ressources humaines capables d'apporter des solutions à de telles préoccupations. C'est dans les soucis d'apporter des solutions à cette problématique que le département de mathématique et informatique de la faculté des sciences et techniques de l'Université Cheikh Anta Diop de Dakar a ouvert depuis 2004 un master de transmission des données et sécurité de l'information.

Ce master a pour objectif d'assurer la formation de cadres ayant des connaissances solides dans les domaines qui concernent les communications, la fiabilité, l'intégrité, le secret de la transmission de l'information, la sécurité des systèmes d'information, l'audit informatique, le développement de solutions de sécurité, les réseaux etc...

Pour l'obtention du master professionnel les étudiants doivent élaborer un mémoire de fin de cycle.

C'est dans cette optique que nous avons rédigé ce document qui s'intitule : Etude et mise en place d'un SIEM pour la gestion de la sécurité.

RESUME :

Ce mémoire de fin d'Etudes traite l'étude et la mise en place d'un SIEM pour la gestion de la sécurité qui est une solution de sécurité aux nouvelles attaques dont sont victimes les entreprises.

La technologie SIEM assure la collecte, le regroupement, la normalisation et la conservation des logs adéquats, l'analyse, la présentation via des rapports et de la visualisation mais aussi des workflows et les contenus relatifs à la sécurité. Tous les cas pratiques d'utilisation de la technologie SIEM se concentrent sur la sécurité des informations, du réseau et des données mais aussi sur la conformité aux règlementations.

Auparavant ils existaient des méthodes de gérer les événements de sécurité d'un système d'information notamment, l'analyse en temps réel de fichiers de journalisation (logs) d'équipements réseau comme l'analyse des logs d'un pare-feu en temps réel afin de détecter une attaque et mettre en place une règle de filtrage adéquate. D'autre part, l'externalisation des logs sur un serveur central afin de conserver les événements de sécurité et les analyser en cas de problème.

Aujourd'hui les SIEM nous permettent de faire converger ces deux fonctionnalités et de fournir une solution complète permettant de collecter, de normaliser, d'agréger, d'archiver, d'analyser, d'alerter, de corréler et de fournir des tableaux de bords des événements de sécurité du système d'information.

ABSTRACT:

This end of studies thesis deals with the study and implementation of a SIEM for security management, which is a security solution to new attacks to which companies are victims.

SIEM technology ensures the collection, aggregation, standardization and retention of adequate logs, analysis, presentation through reports and visualization as well as workflows and security-related content. All SIEM technology use cases focus on information, network and data security as well as regulatory compliance.

Previously there were methods of managing the security events of an information system, in particular, the real-time analysis of log files (logs) of network equipment such as the analysis of logs from a firewall in real time in order to detect an attack and set up an adequate filtering rule. On the other hand, the outsourcing of logs to a central server in order to keep security events and analyze them in the event of a problem.

Today SIEM allows us to converge these two functionalities and to provide a complete solution making it possible to collect, standardize, aggregate, archive, analyze, alert, correlate and provide tables of edges of information system security events.

INTRODUCTION GENERALE :

De nos jours avec l'avancée fulgurante des nouvelles technologies de l'information et de la communication, la multiplication et la diversité des objets connectés et l'utilisation de l'internet en entreprise, indispensable pour échanger 1des informations dans un cadre de travail. Avec la possibilité d'échanger des données avec des interlocuteurs distants et parfois inconnus, la capacité à identifier de façon certaine l'auteur d'un message, d'une transaction ou d'un document devient fondamentale. Cette ascension des nouvelles technologies s'accompagne avec des problèmes liés à la sécurité informatique tels que les hackers, la propagation de virus et de programmes malveillants.

Ce besoin évident de sécurité a engendré le développement de diverses solutions de sécurité comme l'antivirus, les firewalls, les détecteurs d'intrusions, les équipements de supervisions, les VPN et récemment la gestion et l'analyse des logs.

Mais ces solutions bien vraies qu'ils ont toujours joué un rôle capital dans la sécurité et la protection des informations échangées en garantissant l'intégrité, la confidentialité et l'authentification, la non répudiation et la disponibilité des de celles-ci elles restent encore limités dans la gestion des événements de sécurité d'un système d'information.

Tant que cette question de la prise en charge de l'évènement de sécurité reste sans réponse les entreprises et les utilisateurs restent exposés à des attaques.

Auparavant ils existaient des méthodes de gérer les événements de sécurité d'un système d'information notamment, l'analyse en temps réel de fichiers de journalisation (logs) d'équipements réseau comme l'analyse des logs d'un pare-feu en temps réel afin de détecter une attaque et mettre en place une règle de filtrage adéquate. D'autre part, l'externalisation des logs sur un serveur central afin de conserver les événements de sécurité et les analyser en cas de problème.

Aujourd'hui les SIEM nous permettent de faire converger ces deux fonctionnalités et de fournir une solution complète permettant de collecter, de normaliser, d'agréger, d'archiver, d'analyser, d'alerter, de corréler et de fournir des tableaux de bords des événements de sécurité du système d'information.

La protection des données et des informations numériques demeure pour les entreprises un défi permanent à relever. La cryptologie et les techniques de codage de l'information s'appuient sur des concepts mathématiques avancés combinés avec les outils de gestion des

événements et des informations de sécurité peuvent apporter des solutions simples et efficaces à ces problèmes. L'université Cheikh Anta DIOP de Dakar, à travers le Laboratoire d'Algèbre, de Cryptologie et de Géométrie Algébrique et Applications (LACGAA) met sur le marché depuis 2004 des ingénieurs de haut niveau capables de concevoir, développer et mettre en place des solutions de sécurité pour répondre aux besoins des entreprises.

Ce master a pour mission de former des cadres ayant des connaissances solides dans les domaines qui concernent la sécurité de l'information et de la communication : la fiabilité, l'intégrité et le secret de la transmission de l'information, l'Audit informatique, le développement de solutions de sécurité, la sécurité des réseaux, etc.

Chapitre 1 : Cadre théorique et méthodologique :

I. Cadre théorique

I-1-Définition du besoin

Aujourd'hui, la grande majorité des systèmes, des applications et éléments réseau au sein des entreprises produisent des logs. Dans la plupart des cas, ces logs sont stockés de manière locale sur les machines.

Ces données gérées par les administrateurs deviennent de plus en plus nombreuses sans être forcément pertinentes. De ce fait, sans un outil de traitement automatisé, certaines alertes critiques risquent de se retrouver inondées sous la masse d'information.

Dans ce cadre, la centralisation des logs consiste à mettre sur un même système, l'ensemble des logs des systèmes, applications et services des machines du SI. Cette action suit le même principe que la supervision, qui consiste à centraliser les évènements de surveillance sur un système unique.

La gestion des événements et incidents de sécurité de l'information constitue un élément important de la sécurité de l'information. Cette démarche consiste à définir un ensemble de mesures techniques et organisationnelles pour faire face aux différentes menaces qui pèsent sur le patrimoine informationnel d'une structure. Dans cette perspective le SIEM (Security Information and Events Management) permet la collecte et l'agrégation des données d'une multitude de sources (réseau, serveurs, applications, etc.), la corrélation des événements, la détection des événements de sécurité la notification du personnel, la rétention des données, l'élaboration des tableaux de bords etc.

I-2-Importance de la question

L'utilisation du SIEM en entreprise devient indispensable pour assure la sécurité du SI dans le contexte actuel où l'identification des menaces connues et inconnues devient un enjeu majeur pour les entreprises. Il constitue une tour de contrôle des événements de sécurité de l'information pour alimenter le processus de réponse aux incidents de sécurité mais aussi un moyen d'avoir une visibilité sur le manque d'efficacité des contrôles de sécurité mis en place.

En outre, l'intérêt de cette démarche se confirme à travers les nouvelles tendances que connaît le marché des nouvelles technologies en permettant de renforcer le lien entre l'entreprise et son environnement externe (clients, partenaires, etc.) mais aussi son environnement interne

avec ses collaborateurs par le biais de solutions de mobilité qui ont commencé à voir le jour ces derniers mois.

Ainsi, les entreprises ont de plus en plus besoin de mettre en place un moyen pour centraliser la supervision et par conséquent la gestion des événements de la sécurité de l'information. Ceci, afin de faire face au volume exorbitant des événements de sécurité généré par l'introduction de ces nouvelles tendances mais aussi assurer la conformité par rapport aux exigences réglementaires relatives au reporting, à la rétention des logs, à la notification, etc.

I-3-Formulation des objectifs

La sécurité des systèmes d'information est très bien prise en charge en général dans les entreprises. Mais toute politique de sécurité se veut dynamique et évolutive en s'adaptant en permanence aux nouvelles formes d'attaque multiples et subtiles. C'est dans ce cadre que nous avons choisi comme sujet de mémoire l'étude et la mise en place d'une solution SIEM pour la gestion de la sécurité.

Pour la réalisation de cette solution nous allons d'abord faire une étude théorique des différentes solutions SIEM, ensuite faire une étude comparative de ces solutions et en choisir celle que nous jugerons la meilleure.

II. Méthodologie

II-1-Délimitation du sujet :

Ce travail de recherche se concentre sur la gestion centralisée de la Sécurité des Systèmes d'Information, et plus particulièrement sur les SIEM (Security Information and Event Management). Nous allons parler d'abord des différentes solutions de sécurité, voir les solutions de supervision et la gestion des logs. Ensuite nous allons nous intéresser à l'utilité d'une telle gestion en comprenant pourquoi les entreprises en ont besoin de nos jours, puis étudier le fonctionnement théorique d'un tel système, et enfin nous plonger dans quelques mises en pratique. Pour terminer, nous comparerons nos résultats pour déterminer le meilleur moyen de protéger le système de sécurité d'une entreprise.

II-2-Les techniques d'investigation :

Le thème Etude et mise en place d'une solution SIEM pour la gestion de la sécurité développé dans ce document est le résultat des recherches menées autour de l'actualité encadrant la sécurité informatique en la gestion des informations et événements de sécurité à travers les

technologies SIEMs. En effet, ce travail de recherche nous a permis d'explorer le paysage dynamique de la sécurité informatique et la place des technologies SIEMs dans ce dernier.

Notre étude organisée en huit chapitres à savoir :

- Le chapitre1 qui étudie la problématique, les objectifs que nous rechercherons, la solution que nous proposons et les raisons qui nous ont poussés à choisir ce sujet,
- Le chapitre 2 où nous dégageons la méthodologie à suivre
- Le chapitre 3 aborde la sécurité des SI dans sa globalité
- Le chapitre 4 étudie les différentes solutions de sécurité
- Le chapitre 5 parle de l'étude théorique des SIEM
- Le chapitre 6 le SIEM comme solution de sécurité
- Le chapitre 7 détaille le déploiement de Splunk ES
- Et en fin dans le chapitre8 nous avons fait des tests sur quelques domaines d'application de Splunk ES.

II-3-Echantillonnage :

Dans la mise en œuvre nous avons étudié un cas de système basique adaptable à la plupart des systèmes en entreprise. Ce dernier a été déployé et testé dans un environnement virtuel à l'aide des outils VirtualBox et Gns3 suivant le plan d'adressage et les équipements ci-dessous :

Poste	**SE et Version**	**Adresse IP**	**Masque**	**Passerelle**	**Interfaces**
Host	ElementaryOS Hera 5.1	172.16.31.1	255.255.255.0		vboxnet0
		DHCP	DHCP	DHCP	wl0
IPCFW	IPCop 2.1.8	172.16.31.2	255.255.255.0	172.16.31.1	rouge
		10.0.1.1	255.255.255.0		vert
		10.0.2.1	255.255.255.0		orange
WEB-SRVR	Ubuntu server 18.04	10.0.2.20	255.255.255.0	10.0.2.1	enp0s3
Splunk-Indexer	Ubuntu server 20.04	10.0.2.30	255.255.255.0	10.0.2.1	enp0s3

II-4-Les difficultés rencontrées :

Durant ce travail nous nous sommes confrontés à plusieurs difficultés à savoir :

Le manque de documentation sur les SIEM

L'absence d'un environnement permettant de faire des tests approfondis

Les limites de la virtualisation

Les limites de la licence essais de Splunk

Le manque de formation dédiée sur Splunk

Splunk étant un logiciel à fonctionnalités multiples, nous nous sommes limités à présenter l'essentiel et faire ressortir son rôle dans la gestion de la sécurité d'un système d'information.

Chapitre 2 : Sécurité des systèmes d'information

III. Généralités sur la sécurité

L'échange d'informations en entreprises, nécessite un réseau sécurisé pour le transfert des données aussi bien entre les machines en local qu'avec les machines externes. Cela étant, la sécurité de façon générale est présente à plusieurs niveaux et elle est à prendre dans sa totale dimension : données, applications, équipements et trafic réseau.

La continuité de l'activité de l'entreprise appelle celle de son système d'information.

Cette continuité ne peut être assurée que par la mise en place de moyens de protection apportant un niveau de sécurité adapté aux enjeux spécifiques de l'entreprise.

La sécurité des réseaux est devenue l'un des éléments clés de la continuité des SI de l'entreprise. Comme toute composante critique, le réseau doit faire l'objet d'une politique de sécurité tenant compte de tous les besoins d'accès au réseau d'entreprise :

- Accès distants
- Echange des mails
- Commerce électronique
- Interconnexion des tierces parties etc.

La sécurité informatique a cinq objectifs à savoir :

- ✓ **L'intégrité :** c'est-à-dire garantir que les données sont bien celles que l'on croit être.
- ✓ **La confidentialité** : consistant à assurer que seules les personnes autorisées aient accès aux ressources échangées.
- ✓ **La disponibilité** : les services (ordinateurs, réseaux, périphériques, applications…) et les informations (données, fichiers…) doivent être accessibles aux personnes autorisées quand elles en ont besoin.
- ✓ **La non-répudiation :** permettant de garantir qu'une transaction ne peut être niée.
- ✓ **L'authentification :** consistant à s'assurer de l'identité des entités.

La sécurité regroupe deux notions distinctes à savoir la capacité à lutter contre des problèmes accidentels et naturels mais aussi de lutter contre une action humaine malveillante.

Si le premier problème est avant tout un problème d'infrastructure.

Le deuxième problème requiert non seulement des mesures, mais fait également appel à de nombreuses autres techniques.

III-1-Les différents aspects de la sécurité

La sécurité doit être abordée dans un contexte global et notamment prendre en compte les aspects suivants :

La sécurité physique, soit la sécurité au niveau des infrastructures matérielles : salles sécurisées, lieux ouverts au public, espaces communs de l'entreprise, postes de travail des personnels.

La sécurité personnelle : la sensibilisation des utilisateurs aux problèmes de sécurité.

La sécurité logique : c'est-à-dire la sécurité au niveau des données, notamment les données de l'entreprise, les applications ou encore les systèmes d'exploitation.

La sécurité des communications : technologies réseau, serveurs de l'entreprise, réseaux d'accès, etc.

La sécurité procédurale : sert de tampon entre les commandes juridiques et les livraisons technique.

III-2-La politique de sécurité

La Politique de Sécurité du Système d'Information définit l'intégralité de la stratégie de sécurité informatique de l'entreprise. Elle se traduit par la réalisation d'un document qui regroupe l'ensemble des règles de sécurité à adopter ainsi que le plan d'actions ayant pour objectif de maintenir le niveau de sécurité de l'information dans l'entreprise.

Elle décrit l'ensemble des enjeux, des besoins, des contraintes, ainsi que des règles à adopter propres à chaque structure. Elle doit être validée par la direction et prise en compte par chaque collaborateur.

En premier lieu, la démarche de mise en œuvre d'une PSSI permet d'entreprendre une évaluation de la maturité de la sécurité de l'organisme, d'identifier les failles et les faiblesses organisationnelles et techniques afin de prévoir et d'appliquer un plan d'actions correctives et des règles associées.

La Politique de Sécurité du Système d'Information est un document qui doit être pris en compte par chaque collaborateur intervenant dans l'organisme afin qu'il puisse connaître les règles et les enjeux en matière de sécurité interne et externe mais aussi des mesures à appliquer en situation de crise liée.

De surcroît définir une politique de sécurité permet également d'évaluer l'importance du rôle que joue le système d'information dans le fonctionnement de l'ensemble des services.

L'élaboration d'une politique de sécurité nécessite une approche globale, portant à la fois sur l'aspect technique tel que la sécurité logique, la sécurité informatique et la sécurité des réseaux mais également la sécurité physique, organisationnelle ainsi que les aspects liés à l'humain.

Cette conception ne peut s'effectuer seul. Il s'agit d'un travail d'équipe piloté la plupart du temps par le Responsable de la Sécurité du Système d'Information avec l'aide de la Direction et de chaque personne composant la gouvernance de la sécurité de l'information mais aussi d'autres collaborateurs tels que certains responsables de services et des ressources humaines.
Ensemble, le groupe de travail effectuera une analyse du niveau de maturité SSI permettra de définir le périmètre de la politique de sécurité ainsi que ses objectifs.
Ensuite, un minimum des domaines ci-dessous devra constituer la Politique de Sécurité du Système d'Information :

- Principes organisationnels
- Politique de sécurité
- Organisation de la sécurité
- Gestion des risques SSI
- Sécurité et cycle de vie
- Assurance et certification
- Principes de mise en œuvre
- Aspects humains
- Planification de la continuité des activités
- Gestion des incidents
- Sensibilisation et formation
- Exploitation
- Aspects physiques et environnementaux
- Principes techniques
- Identification / authentification
- Contrôle d'accès logique
- Journalisation
- Infrastructures de gestion des clés cryptographiques
- Signaux compromettants.

La mise en œuvre d'une Politique de Sécurité du Système d'Information nécessite une méthodologie particulière afin d'impliquer chaque acteur concerné et prendre en compte l'ensemble des domaines nécessaires à sa construction.

IV. Différentes solutions de sécurité :

IV.1 Les Firewall (Pare-feu) :

Un firewall aussi appelé pare-feu est un matériel ou un logiciel qui vise à filtrer les informations qui entrent et qui sortent d'un réseau informatique. C'est un des éléments indispensables au maintien de la sécurité de données. Cet outil prévient différentes menaces qui pourraient nuire à l'entreprise. En d'autres termes, il protège des intrusions provenant d'un réseau extérieur et surtout d'internet.

Chaque ordinateur connecté à internet est susceptible d'être victime d'une attaque d'un pirate informatique. La méthodologie généralement employée par le pirate informatique consiste à scruter le réseau à la recherche d'une machine connectée, puis à chercher une faille de sécurité afin de l'exploiter et d'accéder aux données s'y trouvant.

Cette menace est d'autant plus grande que la machine est connectée en permanence à internet pour plusieurs raisons :

- La machine cible est susceptible d'être connectée sans pour autant être surveillée ;
- La machine cible est généralement connectée avec une plus large bande passante ;
- La machine cible ne change pas (ou peu) d'adresse IP.

Ainsi, il est nécessaire, autant pour les réseaux d'entreprises que pour les internautes possédant une connexion de type câble ou ADSL, de se protéger des intrusions réseaux en installant un dispositif de protection.

Le pare-feu est un système permettant de filtrer les paquets de données échangés avec le réseau, il s'agit ainsi d'une passerelle filtrante comportant au minimum les interfaces réseau suivante :

- une interface pour le réseau à protéger (réseau interne) ;
- une interface pour le réseau externe.

Le système firewall est un système logiciel, reposant parfois sur un matériel réseau dédié, constituant un intermédiaire entre le réseau local (ou la machine locale) et un ou plusieurs réseaux externes. Il est possible de mettre un système pare-feu sur n'importe quelle machine et avec n'importe quel système pourvu que :

- La machine soit suffisamment puissante pour traiter le trafic ;
- Le système soit sécurisé ;
- Aucun autre service que le service de filtrage de paquets ne fonctionne sur le serveur. Dans le cas où le système pare-feu est fourni dans une boîte noire « clé en main », on utilise le terme d'« Appliance ».

a- Fonctionnement d'un pare-feu :

Un système pare-feu contient un ensemble de règles prédéfinies permettant :

- D'autoriser la connexion ;
- De bloquer la connexion ;
- De rejeter la demande de connexion sans avertir l'émetteur.

L'ensemble de ces règles permet de mettre en œuvre une méthode de filtrage dépendant de la politique de sécurité adoptée par l'entité. On distingue habituellement deux types de politiques de sécurité permettant :

- Soit d'autoriser uniquement les communications ayant été explicitement autorisées,
- Soit d'empêcher les échanges qui ont été explicitement interdits.

La première méthode est sans nul doute la plus sûre, mais elle impose toutefois une définition précise et contraignante des besoins en communication.

b- Le filtrage simple de paquets :

Un système pare-feu fonctionne sur le principe du filtrage simple de paquets. Il analyse les en-têtes de chaque paquet de données (datagramme) échangé entre une machine du réseau interne et une machine extérieure.

Ainsi, les paquets de données échangées entre une machine du réseau extérieur et une machine du réseau interne transitent par le pare-feu et possèdent les en-têtes suivants, systématiquement analysés par le firewall :

- Adresse IP de la machine émettrice ;
- Adresse IP de la machine réceptrice ;
- Type de paquet (TCP, UDP, etc.) ;
- Numéro de port (rappel : un port est un numéro associé à un service ou une application réseau).

Les adresses IP contenues dans les paquets permettent d'identifier la machine émettrice et la machine cible, tandis que le type de paquet et le numéro de port donnent une indication sur le type de service utilisé.

Les ports reconnus (dont le numéro est compris entre 0 et 1023) sont associés à des services courants (les ports 25 et 110 sont par exemple associés au courrier électronique, et le port 80 au Web). La plupart des dispositifs pare-feu sont au minimum configuré de manière à filtrer les communications selon le port utilisé. Il est généralement conseillé de bloquer tous les ports qui ne sont pas indispensables.

Le port 23 est par exemple souvent bloqué par défaut par les dispositifs pare-feu car il correspond au protocole Telnet, permettant d'émuler un accès par terminal à une machine distante de manière à pouvoir exécuter des commandes à distance. Les données échangées par Telnet ne sont pas chiffrées, ce qui signifie qu'un individu est susceptible d'écouter le réseau et de voler les éventuels mots de passe circulant en clair. Il est préférable d'utiliser le protocole SSH, réputé sûr et fournissant les mêmes fonctionnalités que Telnet.

c- Le filtrage dynamique

Le filtrage simple de paquets ne s'attache qu'à examiner les paquets IP indépendamment les uns des autres, ce qui correspond au niveau 3 du modèle OSI. Or, la plupart des connexions reposent sur le protocole TCP, qui gère la notion de session, afin d'assurer le bon déroulement des échanges. D'autre part, de nombreux services (le FTP par exemple) initient une connexion sur un port statique, mais ouvrent dynamiquement (c'est-à-dire de manière aléatoire) un port afin d'établir une session entre la machine faisant office de serveur et la machine cliente.

Ainsi, il est impossible avec un filtrage simple de paquets de prévoir les ports à laisser passer ou à interdire. Pour y remédier, le système de filtrage dynamique de paquets est basé sur l'inspection des couches 3 et 4 du modèle OSI, permettant d'effectuer un suivi des transactions entre le client et le serveur. Le terme anglo-saxon est « stateful inspection » ou « stateful packet filtering », traduisez « filtrage de paquets avec état ».

Un dispositif pare-feu de type « stateful inspection » est ainsi capable d'assurer un suivi des échanges, c'est-à-dire de tenir compte de l'état des anciens paquets pour appliquer les règles de filtrage. De cette manière, à partir du moment où une machine autorisée initie une connexion avec une machine située de l'autre côté du pare-feu ; l'ensemble des paquets transitant dans le cadre de cette connexion seront implicitement acceptés par le pare-feu.

Si le filtrage dynamique est plus performant que le filtrage de paquets basique, il ne protège pas pour autant de l'exploitation des failles applicatives liées aux vulnérabilités des

applications. Or ces vulnérabilités représentent la part la plus importante des risques en termes de sécurité.

d- Le filtrage applicatif :

Le filtrage applicatif permet de filtrer les communications application par application. Le filtrage applicatif opère donc au niveau 7 du modèle OSI, contrairement au filtrage de paquets simple. Le filtrage applicatif suppose donc une connaissance des protocoles utilisés par chaque application.

Le filtrage applicatif permet, comme son nom l'indique, de filtrer les communications application par application. Le filtrage applicatif suppose donc une bonne connaissance des applications présentes sur le réseau, et notamment de la manière dont elle structure les données échangées.

Un firewall effectuant un filtrage applicatif est appelé généralement « passerelle applicative » (ou « proxy »), car il sert de relais entre deux réseaux en s'interposant et en effectuant une validation fine du contenu des paquets échangés. Le proxy représente donc un intermédiaire entre les machines du réseau interne et le réseau externe, subissant les attaques à leur place. De plus, le filtrage applicatif permet la destruction des en-têtes précédant le message applicatif, ce qui permet de fournir un niveau de sécurité supplémentaire.

Il s'agit d'un dispositif performant, assurant une bonne protection du réseau, pour peu qu'il soit correctement administré. En contrepartie, une analyse fine des données applicatives requiert une grande puissance de calcul et se traduit donc souvent par un ralentissement des communications, chaque paquet devant être finement analysé.

Par ailleurs, le proxy doit nécessairement être en mesure d'interpréter une vaste gamme de protocoles et de connaître les failles afférentes pour être efficace.

Enfin, un tel système peut potentiellement comporter une vulnérabilité dans la mesure où il interprète les requêtes qui transitent par son biais. Ainsi, il est recommandé de dissocier le pare-feu (dynamique ou non) du proxy, afin de limiter les risques de compromission.

e- Les limites des pare-feu :

Un système pare-feu n'offre bien évidemment pas une sécurité absolue, bien au contraire. Les firewalls n'offrent une protection que dans la mesure où l'ensemble des communications vers l'extérieur passe systématiquement par leur intermédiaire et qu'ils sont correctement

configurés. Ainsi, les accès au réseau extérieur par contournement du firewall sont autant de failles de sécurité. C'est notamment le cas des connexions effectuées à partir du réseau interne à l'aide d'un modem ou de tout moyen de connexion échappant au contrôle du pare-feu.

De la même manière, l'introduction de supports de stockage provenant de l'extérieur sur des machines internes au réseau ou bien d'ordinateurs portables peut porter fortement préjudice à la politique de sécurité globale.

Enfin, afin de garantir un niveau de protection maximal, il est nécessaire d'administrer le pare-feu et notamment de surveiller son journal d'activité afin d'être en mesure de détecter les tentatives d'intrusion et les anomalies. Par ailleurs, il est recommandé d'effectuer une veille de sécurité afin de modifier le paramétrage de son dispositif en fonction de la publication des alertes.

La mise en place d'un firewall doit donc se faire en accord avec une véritable politique de sécurité.

IV.2 Les systèmes de détection et de prévention d'intrusion :

a- IDS :

La détection des intrusions est le processus de surveillance des événements se trouvant dans un système des ordinateurs ou du réseau en les analysant pour détecter les signes des intrusions, définies comme des tentatives pour compromettre la confidentialité, l'intégrité, la disponibilité ou éviter des mécanismes de sécurité de l'ordinateur ou du réseau. L'intrusion est causée par les attaques accédant au système via l'Internet, autorisée l'utilisateur du système qui essaye à gagner les privilèges supplémentaires pour lesquels ils ne sont pas autorisés, et autoriser les utilisateurs qui abusent des privilèges donnés. Le système de détection des intrusions est un logiciel ou un matériel qui automatise des surveillances et les processus analysés.

Rôle des IDS :

Les IDS ont pour principal rôle de surveiller la circulation des paquets sur le réseau. Quand une tentative est réussie en passant votre par feu, il va peut-être provoquer des menaces. Alors, vous pouvez diminuer des fautes positives en connaissant ces tentatives. Dans l'environnement de NAT est aussi un profit parce qu'il nous permet de tenir l'adresse réelle de la source et la mettre en corrélation avec des événements entre le système IDS qui se situe avant et après le par feu. Il permettra de vérifier que la ligne de base du par feu est suivi, ou

que quelqu'un a fait une erreur en changeant une règle de par feu. Si nous savons par exemple que la ligne de base du par feu proscrit l'utilisation de ftp et que le système IDS montre des alertes de ftp, alors nous saurons que le par feu ne bloque pas de trafic de ftp. C'est juste un effet secondaire et ne devrait pas être la seule manière de vérifier la conformité à la ligne de base.

Les types d'IDS :

Nous avons plusieurs types d'IDS disponibles de nos jours qui sont caractérisés par des surveillances différentes et des approches d'analyse. Chaque approche a toujours des avantages et des inconvénients. De plus, toute approche peut être décrite dans un terme du model de processus généraux pour IDS.

Network-Based IDSs (NIDS) :

Un système de détection d'intrusion basé sur le réseau est conçu pour aider les organisations à surveiller leurs environnements Cloud, sur site et hybrides pour détecter les événements suspects qui pourraient indiquer un compromis. Cela inclut les violations de politique et l'analyse des ports, ainsi que le trafic source et de destination inconnus.

Les technologies de sécurité NIDS sont de nature « passive » plutôt que « active ». Cela signifie qu'ils sont conçus uniquement pour alerter sur les activités suspectes et, pour cette raison, sont souvent déployés aux côtés de systèmes de prévention des intrusions (IPS) qui sont «actifs ». Pour les organisations qui cherchent à accroître davantage la visibilité des menaces, les systèmes NIDS sont couramment utilisés en conjonction avec les systèmes de détection d'intrusion basés sur l'hôte (HIDS) et les solutions SIEM, qui regroupent et analysent les événements de sécurité de plusieurs sources. Compte tenu de la nature avancée du cyber menaces actuelles, la protection de l'entreprise nécessite désormais la capacité de détecter et de répondre aux attaques qui contournent les pare-feu et autres contrôles de sécurité périmétriques traditionnels. Les systèmes de détection d'intrusion basés sur le réseau (NIDS) aident à améliorer la visibilité des activités malveillantes à l'intérieur du réseau qui pourraient autrement passer inaperçues.

Avantage:

- Le NIDS peut surveiller un grand réseau.
- Le déploiement de NIDS a peu d'impact sur un réseau existant. Les NIDS sont habituellement des dispositifs passifs qui écoutent sur un fil de réseau sans interférer

l'opération normale d'un réseau. Ainsi, il est habituellement facile de monter en rattrapage un réseau pour inclure IDS avec l'effort minimal.

Inconvénients:

- Il est difficile à traiter tous les paquets circulant sur un grand réseau. De plus il ne peut pas reconnaître des attaques pendant le temps de haut trafic.
- Plusieurs des avantages de NIDS ne peuvent pas être appliqués pour les commutateurs modernes. La plupart des commutateurs ne fournissent pas des surveillances universelles des ports et limitent la gamme de surveillance de NIDS. Même lorsque les commutateurs fournissent de tels ports de surveillance, souvent le port simple ne peut pas refléter tout le trafic traversant le commutateur.
- Le NIDS ne peut pas analyser des informations chiffrées (cryptées). Ce problème a lieu dans les organisations utilisant le VPN.
- La plupart de NIDS ne peuvent pas indiquer si une attaque est réussie ou non. Il reconnaît seulement qu'une attaque est initialisée. C'est-à-dire qu'après le NIDS détecte une attaque, l'administrateur doit examiner manuellement chaque host s'il a été en effet pénétré.
- Quelques NIDS provoquent des paquets en fragments. Ces paquets mal formés font devenir l'IDS instable.

Host Based IDS (HIDS):

Les cyber menaces étant désormais plus répandues que jamais, il est essentiel de pouvoir détecter les attaques qui contournent la sécurité du périmètre. Les systèmes de détection d'intrusion basés sur l'hôte (HIDS) aident les entreprises à identifier les menaces à l'intérieur du périmètre du réseau en surveillant les périphériques hôtes pour détecter toute activité malveillante qui, si elle n'est pas détectée, pourrait entraîner de graves violations.

Les systèmes de détection d'intrusion basés sur l'hôte aident les entreprises à surveiller les processus et les applications s'exécutant sur des périphériques tels que des serveurs et des postes de travail. Le HIDS suit les modifications apportées aux paramètres de registre et à la configuration critique du système, aux fichiers journaux et de contenu, alertant en cas d'activité non autorisée ou anormale. Les technologies HIDS sont de nature « passive», ce qui signifie que leur objectif est d'identifier les activités suspectes et non de les empêcher. Pour

cette raison, les solutions HIDS sont souvent utilisées en conjonction avec des systèmes de prévention des intrusions (IPS), qui sont «actifs». Pour les organisations qui souhaitent obtenir une meilleure visibilité de la sécurité, les systèmes de détection d'intrusion basés sur l'hôte sont généralement déployés avec les systèmes de détection d'intrusion basés sur le réseau (NIDS) et les solutions SIEM, qui regroupent et analysent les événements de sécurité provenant de plusieurs sources. Pour détecter les menaces, les systèmes de détection d'intrusion basés sur l'hôte nécessitent l'installation de capteurs appelés «agents HIDS» sur des actifs contrôlables. Un système HIDS utilise une combinaison de méthodes de détection basées sur des signatures et basées sur des anomalies. La détection basée sur les signatures compare les fichiers à une base de données de signatures connues pour être malveillantes. La détection basée sur les anomalies analyse les événements par rapport à une ligne de base de comportement système «typique».

Avantage :

- Pouvoir surveiller des événements locaux jusqu'au host, détecter des attaques qui ne sont pas vues par NIDS,
- Marcher dans un environnement dans lequel le trafic de réseau est encrypté, lorsque les sources des informations d'host-based sont générées avant l'encrypte des données ou après le décrypte des données au host de la destination,
- HIDS n'est pas atteint par le réseau commuté.

Inconvénients

- HIDS est difficile à gérer, et des informations doivent configurées et gérées pour chaque host surveillé.
- Puisque au moins des sources de l'information pour HIDS se résident sur l'host de la destination par les attaques, l'IDS peut être attaqué et neutralisé comme une partie de l'attaque,
- HIDS n'est pas bon pour le balayage de réseau de la détection ou l'autre tel que la surveillance qui s'adresse au réseau entier parce que le HIDS ne voit que les paquets du réseau reçus par ses hosts,
- HIDS peut être neutralisé par certaines attaques DoS.

b- Les IPS :

Définition :

Un système de prévention des intrusions (IPS) est une forme de sécurité de réseau qui sert à détecter et prévenir les menaces identifiées. Les systèmes de prévention des intrusions surveillent en permanence le réseau, recherchant les éventuels actes de malveillance et capturent des informations à leur sujet. L'IPS signale ces événements aux administrateurs du système et prend des mesures préventives, telles que la fermeture des points d'accès et la reconfiguration des firewalls pour empêcher de futures attaques. Les solutions IPS peuvent également être utilisées pour identifier les problèmes liés aux politiques de sécurité de l'entreprise, afin de dissuader les employés et les inviter du réseau d'enfreindre les règles incluses dans ces politiques.

Un réseau d'entreprise typique ayant une multitude de points d'accès, il est essentiel de disposer d'un moyen de surveiller les signes d'effraction potentielle, d'incidents et de menaces imminentes. Les menaces réseau actuelles sont de plus en plus sophistiquées et capables d'infiltrer toutes les solutions de sécurité, même les plus robustes.

Principe de fonctionnement des IPS :

Les systèmes de prévention des intrusions fonctionnent en analysant tout le trafic du réseau. Il existe un certain nombre de menaces différentes que les IPS peuvent empêcher, notamment : attaque par déni de service (DDoS) :

- Attaque par déni de service distribué,
- Divers types de failles de sécurité,
- Ver,
- Virus.

L'IPS effectue une inspection des paquets en temps réel, en examinant minutieusement chaque paquet qui circule sur le réseau. Si des paquets malveillants ou suspects sont détectés, l'IPS effectuera l'une des actions suivantes :

- Mettra fin à la session TCP qui a été exploitée et bloquer l'adresse IP source ou le compte utilisateur fautif pour empêcher l'accès non éthique à toute application, hôte cible ou ressource réseau,
- Reprogrammera ou reconfigurera le firewall pour éviter qu'une attaque similaire ne se reproduise à l'avenir,

- Supprimera ou remplacera tout contenu malveillant resté sur le réseau suite à une attaque. Cela est effectué en reconditionnant les charges, en supprimant les informations d'en-tête et en retirant toute pièce jointe infectée des serveurs de fichiers ou de courrier électronique.

Un système de prévention des intrusions est généralement configuré pour utiliser un certain nombre d'approches différentes pour protéger le réseau contre les accès non autorisés. On peut citer :

- Approche basée sur les signatures : cette approche utilise des signatures prédéfinies de menaces réseau bien connues. Lorsqu'une attaque correspondant à l'une de ces signatures ou à l'un de ces modèles est lancée, le système prend les mesures nécessaires.

- Approche basée sur les anomalies : permet de surveiller tout comportement anormal ou inattendu sur le réseau. Si une anomalie est détectée, le système bloque immédiatement l'accès à l'hôte cible.

- Approche basée sur les politiques : cette approche nécessite que les administrateurs configurent les politiques de sécurité en fonction des politiques de sécurité organisationnelles et de l'infrastructure du réseau. Lorsqu'une activité viole une politique de sécurité, une alerte est déclenchée et envoyée aux administrateurs système.

IV.3 La supervision du réseau

La supervision consiste à utiliser les ressources informatiques pour obtenir des informations sur l'état des réseaux et de leurs composants. Ces données seront ensuite traitées et affichées afin de mettre en lumière d'éventuels problèmes.

La supervision peut résoudre les problèmes automatiquement ou dans le cas contraire prévenir les administrateurs. Plusieurs actions sont ainsi réalisées :

- Acquisition de données,
- Analyse, puis
- Visualisation et
- Réaction.

Un tel processus est réalisé à plusieurs niveaux d'un parc de machines :

- Au niveau interconnexions (Réseau),
- Au niveau de la machine elle-même (Système) et
- Au niveau des services offerts par cette machine (Applications).

La supervision réseau fait référence à la surveillance du bon fonctionnement des réseaux informatiques et des services informatiques connectés sur ces réseaux. La surveillance du réseau porte plus spécifiquement sur :

- La qualité (bande passante),
- La sécurité de la connexion Internet, mais aussi, par extension, à
- L'état des services et matériels connectés : serveurs, imprimantes, postes de travail, etc.

La supervision du système se limite à la machine elle-même et en particulier ses ressources. Si l'on souhaite par exemple contrôler la mémoire utilisée ou la charge processeur sur le serveur il va falloir analyser les fichiers de logs système.

La supervision applicative nous permet de vérifier le fonctionnement d'une application lancée sur une machine. Cela peut être par exemple une tentative de connexion sur le port de l'application pour voir si elle retourne ou demande bien les bonnes informations, mais aussi de l'analyse de logs applicatifs.

IV.4 La gestion des Logs

L'analyse des fichiers log est l'évaluation d'un ensemble d'informations enregistrées à partir d'un ou plusieurs événements intervenus sur un environnement IT. Cette pratique peut être utilisée pour :

- Analyser le comportement utilisateur et identifier des modèles de comportement,
- Identifier et anticiper des incidents,
- Etre conforme à la réglementation en place,
- Anticiper et planifier les capacités de l'environnement IT.

Analyser des logs est un véritable challenge et demande un travail fastidieux pour les équipes IT en raison de la volumétrie, mais aussi de la diversité des types de logs, ainsi que les formats propriétaires, les architectures élastiques, etc.

Utiliser un logiciel d'analyse de logs qui exploite des algorithmes de Machine Learning réduit considérablement la charge de travail des équipes IT qui peuvent se concentrer sur des tâches

à valeur ajoutée. Ce logiciel d'analyse de logs permet de surveiller, d'agréger, d'indexer et d'analyser toutes les données des journaux des applications et de l'infrastructure.
Le logiciel d'analyse de logs recueille les évènements tels l'installation d'une application, la tentative de connexion au système, la configuration etc…
Une entrée de log contient des informations telles que :

- La date et l'heure de l'évènement,
- Sur quel support IT s'est produit l'évènement,
- L'identification de l'utilisateur,
- La catégorie : configuration, sécurité,
- Programme à l'origine de l'évènement.

Le logiciel d'analyse de log doit exploiter des algorithmes de Machine Learning afin de réaliser des analyses rapides et précises difficilement réalisables par l'être humain. Ces analyses sont indispensables pour découvrir des anomalies, des divergences, des corrélations, des tendances et établir des prédictions.

Travailler avec des modèles rend l'exploration des journaux plus efficaces et peut aider à obtenir la racine du problème plus rapidement. Analyser les logs et détecter les problèmes devient alors une démarche rapide et pertinente grâce aux algorithmes de Machine Learning.

En un clic, plusieurs millions de lignes de logs insignifiantes sont regroupées en quelques lignes pertinentes en utilisant Oracle Log Analytics par exemple.

Les utilisateurs peuvent ainsi rechercher, explorer et corréler les données de manière à dépanner les problèmes plus rapidement, d'obtenir des informations opérationnelles et de prendre de meilleures décisions.

Chapitre 3 : L'utilisation des outils SIEM pour la sécurité d'un réseau :

V. Généralité sur les SIEMs

V.1 Définition d'un SIEM

Les différents équipements réseaux (pare-feu, switch, routeur, applications, base de données, serveurs…) stockent de plus en plus d'historique d'événements plus souvent appelés logs. Ces logs permettent d'analyser presque en temps réel l'activité d'un processus. Les logs sont donc principalement utilisés pour tenter d'identifier les raisons et les origines d'une panne d'un équipement.

Un système SEM (Security Event Management) centralise le stockage et l'interprétation des logs, et permet une analyse en quasi-temps réel. Le personnel de sécurité peut ainsi mieux s'organiser pour prendre des mesures défensives plus rapidement.

Un système SIM (Security Information Management) collecte des données et les place dans un référentiel central à des fins d'analyse de tendances. Il est ensuite possible de générer des rapports centralisés et automatisés.

Ainsi, un SIEM (Security Information and Event Management) collecte tout document lié à la sécurité, essentiellement les logs (rôle du SEM) pour ensuite les analyser (rôle du SIM).

La gestion des informations et des événements de sécurité (SIEM : Security information and event management) est une solution logicielle qui regroupe et analyse l'activité de nombreuses ressources différentes au sein d'une infrastructure informatique.

Les outils SIEM fonctionnent en rassemblant les données d'événements et de journaux créées par les systèmes hôtes, les applications et les dispositifs de sécurité, tels que les filtres antivirus et les pare-feu, dans toute l'infrastructure d'une entreprise et en rassemblant ces données sur une plate-forme centralisée.

Les outils SIEM identifient et trient les données dans des catégories telles que les connexions réussies ou échouées, les activités de logiciels malveillants et autres activités malveillantes probables.

Ils génèrent ensuite des alertes de sécurité lorsqu'il identifie des problèmes de sécurité potentiels. À l'aide d'un ensemble de règles prédéfinies, les organisations peuvent définir ces alertes comme étant de faible ou de haute priorité.

SIEM collecte les données de sécurité des périphériques suivants :

- Périphériques réseau,
- Serveurs,
- Contrôleurs de domaine,

SIEM stocke, normalise, regroupe et analyse ces données pour découvrir les tendances, détecter les menaces et permettre aux organisations d'enquêter sur les alertes.

Avantage des outils SIEM:

- Ils regroupent les données de plusieurs systèmes,
- Ils les analysent pour détecter les comportements anormaux et les éventuelles cyberattaques,
- Ils collectent les événements et les alertes,
- Ils offrent une vue globale de l'environnement de sécurité dans l'organisation.

Le SIEM est important parce qu'il facilite la gestion de la sécurité pour les entreprises en filtrant des quantités massives de données de sécurité et en hiérarchisant les alertes de sécurité générées par le logiciel.

Le logiciel SIEM permet aux organisations de détecter des incidents qui, autrement, pourraient passer inaperçus.

Le logiciel analyse les entrées du journal pour identifier les signes d'activité malveillante. En outre, comme le système rassemble des événements provenant de différentes sources sur le réseau, il peut recréer la chronologie d'une attaque, permettant ainsi à l'entreprise de déterminer la nature de l'attaque et son impact sur l'activité.

Un système SIEM peut également aider une organisation à répondre aux exigences de conformité en générant automatiquement des rapports qui incluent tous les événements de sécurité enregistrés parmi ses sources. Sans logiciel SIEM, l'entreprise devrait rassembler les données des journaux et compiler les rapports manuellement.

Un système SIEM améliore également la gestion des incidents en permettant à l'équipe de sécurité de découvrir la route qu'emprunte une attaque sur le réseau, identifier les sources qui ont été compromises et fournir les outils automatisés pour prévenir les attaques en cours.

Inconvénients:

- Ils peuvent être coûteux,
- gourmands en ressources,
- parfois compliqués.

- Leur mise en œuvre est longue car elle nécessite un soutien pour assurer une intégration réussie avec les contrôles de sécurité de l'organisation.
- L'analyse, la configuration et l'intégration des rapports requièrent le talent d'experts. C'est pourquoi certains systèmes SIEM sont gérés directement au sein d'un centre d'opérations de sécurité (SOC), une unité centralisée composée d'une équipe de sécurité de l'information et qui s'occupe des questions de sécurité d'une organisation.
- Les outils SIEM dépendent généralement de règles pour analyser toutes les données enregistrées. Le problème, c'est que le réseau d'une entreprise génère un grand nombre d'alertes, généralement 10 000 par jour ce qui peuvent être positives ou non. Difficile, dans ces conditions, d'identifier les attaques potentielles en raison du nombre de journaux non pertinents.
- Pour en finir avec les défauts, il est à savoir qu'un outil SIEM mal configuré peut manquer des événements de sécurité importants, ce qui rend la gestion des risques liés à l'information moins efficace.

V.2 Principe de fonctionnement des SIEM:

Les SIEM utilisent des étapes de récupération, analyse et gestion de l'information, ce sont la collecte, la normalisation, l'agrégation, la corrélation, le reporting et la réponse.

La collecte d'informations :

Les équipements et logiciels de sécurité sont nombreux dans un système d'information, ils gèrent généralement de façon indépendante des informations de sécurité dites «locales». Le principe de l'étape de collecte est de fournir au SIEM des données à traiter. Ces données peuvent être de nature diverse en fonction de l'équipement ou du logiciel, mais aussi être envoyées de manières tout à fait différentes. On distingue deux modes de fonctionnement :

--Mode actif : Le SIEM possède un ou plusieurs agents déployés sur les équipements à superviser. Ces agents ont pour fonction de récupérer les informations des équipements et logiciels de sécurité et de les envoyer au SIEM. Un élément de sécurité qui a été conçu nativement pour être un agent du SIEM est appelé une sonde.

--Mode passif : Le SIEM est en écoute directe sur les équipements à superviser. Pour cette méthode, c'est l'équipement ou le logiciel qui envoie des informations sans intermédiaire au SIEM.

La normalisation des données :

Les informations collectées viennent d'équipements et logiciels hétérogènes ayant pour la plupart leurs propres moyens de formater les données. Cette étape permet d'uniformiser les informations selon un format unique pour faciliter le traitement par le SIEM.

L'agrégation des données :

L'agrégation est le premier traitement des évènements de sécurité. Il consiste en un regroupement d'évènements de sécurité selon certains critères. Ces critères sont généralement définis via des règles appelées règles d'agrégation et s'appliquent à des évènements ayant des similarités. Le rôle principal de l'agrégation est de réduire le nombre d'évènements en associant un «poids» à ceux-ci. Ainsi un regroupement de trois évènements de sécurité selon un critère défini sera un seul évènement avec un poids de trois. Cela facilite notamment le traitement de l'étape de corrélation, qui gère alors non plus des évènements individuels, mais des groupes d'évènements.

La corrélation des données :

La corrélation correspond à l'analyse d'évènements selon certains critères. Ces critères sont généralement définis via des règles appelées règles de corrélation. Le but de cette étape est d'établir des relations entre évènements, pour ensuite pouvoir créer des alertes de corrélations, des incidents de sécurités, des rapports d'activité… La corrélation se différencie sur plusieurs points :

--Auto-apprentissage et connaissances rapportées : Pour pouvoir fonctionner, les moteurs de corrélation ont besoin d'informations sur les systèmes et réseaux de l'infrastructure. Ces informations peuvent être collectées automatiquement et/ou saisies manuellement par un opérateur.

--Temps réel et données retardées : Dans certains cas, les évènements bruts sont forgés et envoyés directement pour être corrélés en temps réel. Dans d'autres cas, les évènements sont d'abord stockés, et envoyés après un premier traitement, leur envoi peut être alors conditionné.

--Corrélation active et passive : La corrélation active a la possibilité de compléter les évènements reçus en recueillant des informations supplémentaires pour prendre des décisions. La corrélation passive est une corrélation qui ne peut pas interagir avec son environnement, elle reçoit des évènements et prend des décisions.

La «cross-corrélation» est une corrélation capable d'associer et de prioriser les évènements de sécurité reçus, mais aussi d'autres informations (scanners de vulnérabilité…). C'est une corrélation active élargie à de nombreux outils.

V.3 Les coûts :

Les coûts d'adoption d'un SIEM dépendent largement de deux facteurs : la robustesse des capacités du SIEM et l'architecture de déploiement. Certains SIEM offrent une solution «light» qui fournit des capacités de gestion de logs et de reporting de base, sans les capacités analytiques avancées que d'autres systèmes proposent. Ces SIEM «light» s'avèrent considérablement moins onéreux à acquérir que les autres. L'architecture de déploiement a également des implications pécuniaires évidentes. La plupart des SIEM nécessitent l'achat de matériels et/ou de logiciels, tandis que les services de SIEM en mode Cloud sont généralement facturés à l'usage.

Dans une étude publiée en 2016, le cabinet d'analystes Gartner estimait à 22 millions de dollars le marché des logiciels de sécurité informatique en 2015 (+3,7 % sur un an) et soulignait que le segment des SIEMs demeurait encore le segment à la plus forte croissance (+15,8 %).

Les recommandations dans l'usage d'un SIEM :

S'affranchir des problèmes de conformité :

La qualité des données analytiques générées par le SIEM dépend des données initiales sur lesquelles le SIEM s'appuie. Les réseaux étant très différents les uns des autres, cela suppose de choisir un outil de log management qui offre une plateforme large, intégrant un grand nombre de sources de logs (les formats Syslog, les fichiers textes simples, les fichiers de bases de données telles que SQL, Oracle, SNMP, etc.).

Extraire les informations de valeur des logs :

L'outil «nourrissant» le SIEM doit être capable de traiter et fournir des données structurées et non-structurées, et doit intégrer des fonctions de transformation comme du filtering, de l'analyse syntaxique, de réécriture, de classification, etc. Grâce à ces fonctions, il ne reste plus qu'à transférer les informations présentant le plus de valeur. Cela permet de réduire significativement les coûts de licences SIEM et de fournir un flux de logs enrichi et reformaté pour une analyse plus efficace. Les cas d'étude concrets estiment à plus de 40 % les économies réalisées sur une année.

Assurer la conformité réglementaire :

Les fonctions de transformation telles que l'anonymisation et le fait de définir des pseudonymes permettent de garantir un respect plus précis aux standards internationaux en matière de respect de la vie privée et de traitement de données, tels que PCI-DSS, HIPAA et la future RGPD au niveau européen.

Compresser les messages de logs :

La bande passante des réseaux intranet et internet variant beaucoup, l'outil de log management doit être capable de travailler dans des situations où la bande passante est très limitée. Compresser les messages de logs à la volée permet de réduire radicalement la consommation de bande passante et d'accélérer la collecte centrale des logs avec pour effet de réagir plus rapidement aux potentiels risques de sécurité et opérationnels.

S'assurer de ne perdre absolument aucune donnée de log :

Perdre un log n'a généralement pas de conséquences directes, mais il peut être le premier et le seul signe d'une fuite de données en cours. Les fonctionnalités de prévention de perte de messages, le contrôle de remise des messages ou des problèmes de destination, sont précieux. Il est finalement important de s'assurer que rien ne déclenche de défaillance temporaire au niveau de l'infrastructure de logging, ou bien que celle-ci n'est tout simplement pas à la hauteur de la tâche à accomplir.

Nourrir le SIEM avec des données de surveillance des utilisateurs privilégiés :

Même si la plupart des activités des utilisateurs laissent des traces dans les logs, plusieurs actions d'utilisateurs (surtout celles exécutées par des utilisateurs privilégiés) ne peuvent être vues dans les logs ou dans les données analytiques fournies par le SIEM. En intégrant son SIEM avec une solution de surveillance des activités privilégiés, l'entreprise peut analyser les activités utilisateurs les plus risquées en temps réel, afin de prévenir les types d'attaques les plus coûteux et les compromissions de comptes privilégiés, souvent à la base d'importantes attaques.

Priorisation des alertes SIEM :

En moyenne, un professionnel de sécurité détient sept minutes par alerte générée par son SIEM pour décider si une attaque est en cours ou si un utilisateur a juste ouvert un email de phishing. Pour optimiser la capacité de détection et de prise de décision, les technologies

d'analyse comportementale peuvent aider à identifier les problèmes de sécurité les plus à risques. Elles se basent sur les privilèges dont bénéficient l'utilisateur ciblé et les différences comportementales observées en temps réel par rapport à une base de références de comportements normaux pour cet utilisateur.

V.4 Les prérequis à la mise en place d'un SIEM :

Nous allons qualifier le scénario le plus courant de «protection SIEM». Parmi les entreprises qui ont déployé des solutions SIEM héritées, beaucoup ont essayé d'envoyer une trop grande quantité de données vers leurs outils SIEM, ce qui a entraîné une surcharge et parfois une perte de fonctions et de données essentielles. Afin de pallier ce problème, une solution de gestion des logs est déployée en amont de leur solution SIEM.

Il est nécessaire d'améliorer sa capacité de réponse avant d'être confronté à une situation nécessitant une réaction plus rapide qu'à l'accoutumée. En effet, il est bien plus facile de se préparer à fournir une réponse que d'assurer un monitoring.

Les solutions de logs management associé avec un outil de SIEM sont multiples et continuent de s'accroître tandis que les scénarios de déploiement des outils SIEM seuls se font plus rares et sont susceptibles de continuer à diminuer.

Toutes entreprises possèdent des logs mais la gestion de ceux-ci diffère d'une société à une autre.

Les sociétés ont besoin d'une solution de gestion des logs dès lors qu'elles en possèdent. Par exemple, si elles doivent réviser des logs provenant d'une seule machine, les outils intégrés au système d'exploitation sont généralement suffisants. Cependant, si leur volume quotidien de logs s'élève à 100 Go (une situation tout à fait réaliste), des outils sophistiqués et, par conséquent, coûteux sont nécessaires.

Dans un rapport de 2009 intitulé «How to Implement SIEM Technology» (Comment mettre en œuvre la technologie SIEM), Gartner recommande vivement de «déployer les fonctions de gestion des logs avant de tenter un déploiement à grande échelle de la gestion des événements en temps réel». En outre, lorsque la technologie SIEM se fonde sur un besoin de conformité, le déploiement doit s'effectuer dans le même ordre : «les premières étapes d'un déploiement SIEM principalement orienté vers la norme PCI consistent à déployer des fonctions de gestion des logs pour les systèmes concernés par l'évaluation PCI». Les entreprises doivent améliorer

leur capacité de réponse avant d'être confrontées à une situation nécessitant une réaction plus rapide qu'à l'accoutumée.

Une entreprise déploie un outil de gestion des logs et commence à l'utiliser de manière efficace afin de garantir sa sécurité et sa conformité mais aussi à des fins d'exploitation. La progression naturelle et logique consiste à déployer un outil SIEM afin de gérer les événements en quasi temps réel.

Les critères requis sont les suivants :

Réponse : Tout d'abord, l'entreprise doit être capable de répondre à des alertes dès leur signalement.

Monitoring : L'entreprise doit posséder ou commencer à développer une fonction de monitoring de la sécurité en mettant en place un centre d'opérations de sécurité ou, tout au moins, une équipe chargée du monitoring périodique.

Personnalisation et réglage : L'entreprise doit accepter d'assumer la responsabilité du réglage et de la personnalisation des outils SIEM déployés. Les déploiements de solutions SIEM prêts à l'emploi sont rarement couronnés de succès ou réussissent peu souvent à atteindre leur plein potentiel.

Personnalisation de l'analyse :

Il est donc nécessaire de mettre tout d'abord en œuvre une méthode d'analyse structurée, de la définition des informations à collecter jusqu'à l'élaboration des tableaux de bord finaux, en passant par la mise en place, nous le verrons, de niveaux de référence.

À titre d'exemple encore, pour un routeur Cisco dont la fonction est de simplement aiguiller le trafic, il y aura peu d'événements à collecter. À l'inverse, une passerelle Internet, proposant de multiples fonctions comme un pare-feu, un composant de filtrage d'URL ou un antivirus, produira en comparaison une multitude de fichiers de journalisation et donc d'évènements à collecter.

De plus, les équipements évoluent constamment ainsi que les logs générés. Il est donc nécessaire d'analyser cette évolution des logs afin d'accroitre les événements considérés comme pertinents pour la sécurité (cas de journaux Windows ou de sondes IDS).

Inventaire des sources :

La première étape consiste à lister l'ensemble des équipements qui constitue les sources des événements ou logs. Les sources seront de plus en plus nombreuses car selon le SANS Institute, une nouvelle génération de SIEM, appelés « SIEM v2 » se concentre sur la collecte la plus vaste possible d'informations, ne se basant plus uniquement sur des informations issues des Syslogs, mais aussi netflow ou SNMP par exemple.

Préparation de l'analyse :

La collecte des données événementielles brutes s'effectue soit par une extraction directe de logs provenant de l'équipement, soit par la réception de ces logs par un agent déjà configuré sur notre environnement.

Le volume de données sera sans doute proportionnel au nombre de fonctionnalités qui sont déployées sur l'équipement.

Munis de ces événements, il sera utile de disposer :

- de la documentation du fournisseur ou éditeur de l'équipement fournissant les logs d'un éditeur de texte permettant de manipuler des chaines via des expressions régulières,
- d'un tableur bureautique (ex : Microsoft Excel).

Analyse syntaxique par élimination :

Les logs ou événements doivent à présent être analysés manuellement. Pour cela, nous allons devoir les classer par catégories :

- les logs non nécessaires ou incomplets,
- les logs relatifs à l'administration de l'équipement,
- les logs relatifs à la disponibilité de l'équipement,
- les logs relatifs aux services assurés par l'équipement (ex : journaux de connexions dans le cas d'un serveur proxy).

Les événements à rejeter et à conserver :

Un certain nombre d'événements sont à rejeter car ils sont :

- Sans lien avec des informations propres à la sécurité,
- insuffisamment explicite,
- non utilisables par des mécanismes de corrélation,

– non utiles pour l'élaboration des rapports finaux.

Une fois les événements collectés et analysés, il reste à les intégrer dans un tableau de bord. Celui-ci sera réalisé en fonction d'un périmètre souhaité.

Par exemple :

– Un tableau de bord pare-feu du réseau interne pour les événements de sécurité associés (leurs logs vont être regroupés ensemble puisqu'ils font face aux mêmes niveaux de menace),

– Un tableau de bord «Authentication, Authorization, Accounting» pourra être construit à partir de l'association de l'ensemble des événements de type «AAA », quels que soient les équipements sources.

Intégration et surveillance :

Au final, l'intégration passe par la création de règles d'analyse syntaxique dans les agents de collecte, la définition des alarmes et la création ou la mise à jour de rapports programmés. Ce travail est accompli pour chaque nouvel équipement ou chaque nouvelle alarme créée.

Il est également important de mettre en œuvre une procédure d'analyses régulières de chaque rapport par les équipes du SOC (Security Operations Center) :

– Analyses quantitatives (nombres d'événements),

– Analyses qualitatives (sévérités des événements).

Les alarmes générées en temps réel mises à part, cette revue hebdomadaire permettra dans un premier temps de supprimer les «bruits» jusqu'à atteindre un niveau de stabilité qui détectera des événements de sécurité peu verbeux.

Chaque évènement non classifié devra également être analysé pour classement éventuel dans les catégories «à rejeter» ou «à conserver». Ces analyses peuvent être aussi longues que le réseau est complexe.

V.5 La sécurité de l'application SIEM :

La sécurisation d'un SIEM est une partie importante lors de sa mise en place. Beaucoup d'entreprises exécutant un SIEM utilisent un agent pour envoyer les logs, particulièrement sur l'environnement Windows. L'agent installé sur les serveurs permet de chiffrer en TLS et de livrer une garantie des logs via le protocole TCP.

Afin de garantir une rapidité d'exécution des serveurs Web pour générer les tableaux de bord, la mise en place du protocole HTTP/2 est nécessaire.

SSL et TLS sont deux protocoles qui permettent d'échanger des données de manière sécurisée.

Ils garantissent :

- L'authentification du serveur,
- La confidentialité des données,
- L'intégrité des données.

Toutes les versions de SSL sont à proscrire. Elles contiennent de très nombreuses vulnérabilités. TLS 1.0 et 1.1 sont à éviter également le plus possible. Aujourd'hui la quasi-totalité des navigateurs et OS modernes sont compatibles avec le TLS 1.2. L'ANSSI recommande également de n'utiliser que TLS.

La mise en place d'une suite cryptographique sur le serveur web (cipher suite en anglais) combine des algorithmes d'échange de clés, d'authentification, de chiffrement par bloc et de génération de code d'authentification de message (MAC). Cela permet d'établir le paramétrage de sécurité pour une connexion réseau utilisant le protocole de communication Secure Socket Layer (SSL) et Transport Layer Security (TLS).

Pour les bases de données, l'étape de chiffrement et déchiffrement pourra être réalisée par la couche d'accès aux données. Ainsi, une attaque de type «Injection SQL» ne donnera accès qu'à des données chiffrées à l'attaquant. Dans le cadre de la certification PCI-DSS, les biens susceptibles d'être chiffrés sont spécifiquement définis.

V.6 Architecture de gestion des informations et des événements de sécurité (SIEM) :

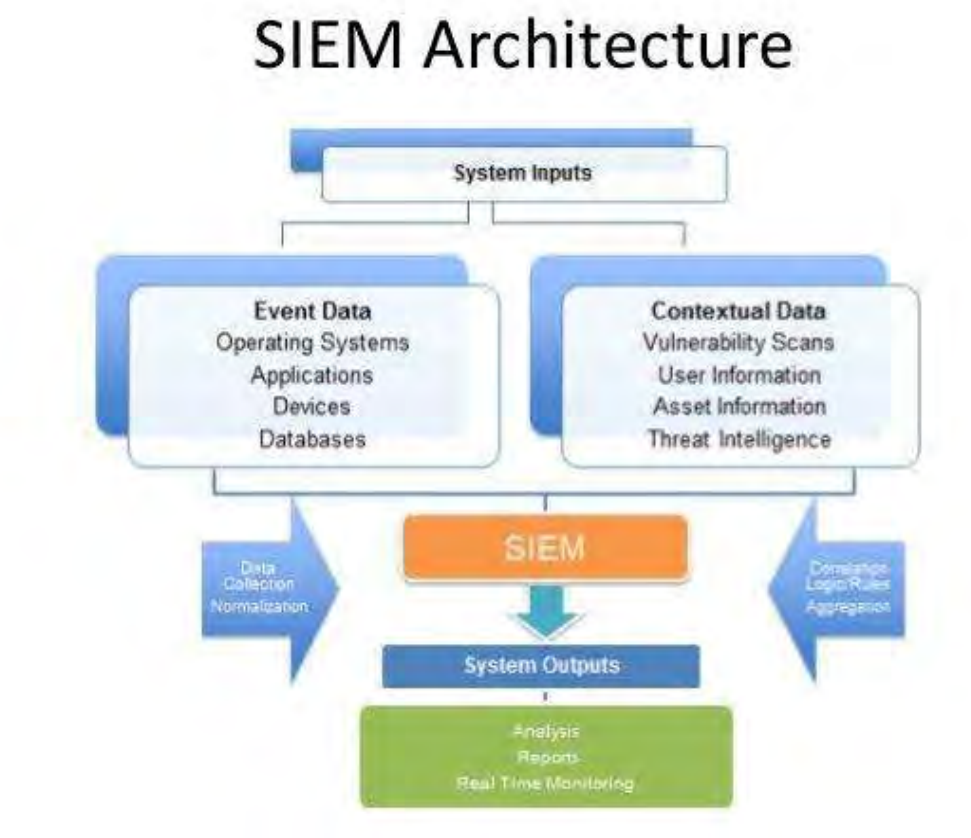


Figure 6.a: Architecture SIEM [W5]

Le SIEM (Security Information and Event Management) présente une large gamme de produits ou de services dans le but de gérer simultanément les informations de sécurité et les événements de sécurité. SIEM fournit également une analyse des alertes de sécurité en temps opportun. D'un point de vue général, le SIEM est utile pour détecter les menaces de sécurité qui ne sont pas visibles par ISS (système de sécurité individuel), enquêter sur les problèmes liés aux violations de sécurité précédentes, effectuer des réponses immédiates aux incidents et préparer des rapports pour répondre aux exigences de conformité.

Afin de faciliter le fonctionnement efficace et complet du SIEM, il faut prêter attention à sa construction, c'est-à-dire à sa technologie et à ses processus architecturaux. De manière aussi précise et concise que possible, cette partie vise à fournir des informations sur le fonctionnement de l'architecture SIEM.

L'un des principaux objectifs de l'architecture SIEM est de maintenir et de gérer les changements de configuration du système, les services d'annuaire, la révision et l'audit des journaux, à la fois les privilèges de service et d'utilisateur avec l'inclusion de la réponse aux incidents. De plus, les applications liées à la gestion des identités et des accès doivent être mis à jour régulièrement pour renforcer la sécurité du système et éliminer les menaces externes. De plus, l'architecture SIEM doit fournir les capacités de présenter, d'analyser et de collecter des informations à partir du réseau et des dispositifs de sécurité. Les fonctions de détection des anomalies et de la visibilité SIEM méritent également d'être mentionnées. La détection du code polymorphe et du zéro jour, l'analyse automatique et la normalisation des journaux peuvent établir des modèles qui sont collectés par la visualisation SIEM en utilisant les événements de sécurité.

L'aspect architectural du SIEM concerne essentiellement le processus de construction des systèmes SIEM et ses composants de base. En un mot, l'architecture SIEM encapsule les composants suivants :

Gestion des journaux : Cela concerne la collecte de données, la gestion des données et la conservation des données précédentes. Le SIEM collecte à la fois les données d'événement et les données contextuelles comme stipulé dans la Figure 6.a. Fondamentalement, l'architecture SIEM collecte des données d'événement à partir de systèmes organisés tels que les périphériques installés, le protocole réseau, les protocoles de stockage (Syslog) et les protocoles de streaming.

La gestion des données concerne principalement les politiques de stockage et de conservation des données. Les SIEM modernes s'appuient sur des technologies offrant des capacités de stockage de données illimitées telles que Hadoop ou Amazon S3. La conservation des données permet de conserver les données pendant une durée spécifique qui est de près de sept ans. Ces données peuvent être utiles à des fins d'analyse judiciaire.

Normalisation des journaux : Il ressort clairement de la figure 6.a que SIEM reçoit l'événement et les données contextuelles en tant qu'entrée. Cependant, la normalisation de celle-ci est nécessaire. Cela concerne la manière dont les données d'événement sont transformées en informations de sécurité pertinentes. Fondamentalement, ce processus implique l'élimination des données non pertinentes, des données générées grâce à un processus de filtrage. L'importance principale de ceci est de ne conserver que les données pertinentes pour une analyse futuriste.

Sources des journaux : les journaux sont collectés à partir des applications réseau, des systèmes de sécurité et des systèmes Cloud. Fondamentalement, ce processus concerne la manière dont les journaux sont introduits dans le SIEM par les organisations.

Choix d'hébergement pour SIEM : Il existe différents modèles disponibles pour l'hébergement du SIEM. Ceux-ci incluent Self-Host, Cloud-Host ou Hybrid-Host.

Reporting de SEIM : Sur la base des journaux disponibles, le SIEM identifie et signale les activités suspectes.

Surveillance en temps réel : SIEM fournit une surveillance en temps réel de l'infrastructure de l'organisation grâce à la détection des menaces et des réponses rapides aux violations de données potentielles.

En conséquence, il est pertinent de souligner que l'architecture SIEM traditionnelle était autrefois monolithique et coûteuse. Cependant, le SIEM de nouvelle génération est plus abordable et offre de meilleurs avantages technologiques grâce à des logiciels sophistiqués et à une technologie Cloud pour une gestion efficace des événements de sécurité.

Il y a plusieurs façons pour un SIEM de gérer des alertes, plusieurs d'entre elles peuvent être utilisés simultanément :

- Le «reporting» : les rapports générés contiennent à la fois une synthèse des alertes et une vue d'ensemble de la sécurité du système à un instant T (statistiques, intrusions, vulnérabilités exploitées, classification des attaques).
- Le stockage : les alertes, incidents et rapports peuvent être stockés dans des bases de données pour pouvoir être analysés ultérieurement par des moteurs de corrélation.
- La réponse : les mécanismes de réponse aux alertes doivent permettre de stopper une attaque ou de limiter ses effets de façon automatique. La réponse à une intrusion dépend de la politique de sécurité.

V.7 Avantages et limites des SIEM :

Avantages des SIEM :

L'utilisation d'un SIEM offre les avantages suivants à l'équipe de sécurité d'une entreprise à savoir :

- Raccourcir le temps nécessaire pour identifier les menaces de manière significative, en minimisant les dommages causés par ces menaces,
- Offrir une vue holistique de l'environnement de sécurité des informations d'une organisation, facilitant la collecte et l'analyse des informations de sécurité pour assurer la sécurité des systèmes,
- Peut être utilisé par les entreprises pour une variété de cas d'utilisation qui tournent autour des données ou des journaux, y compris les programmes de sécurité, les rapports d'audit et de conformité, le service d'assistance et le dépannage du réseau,
- Prend en charge de grandes quantités de données afin que les organisations puissent continuer à évoluer et augmenter leurs données,
- Fournit une détection des menaces et des alertes de sécurité,
- Peut effectuer une analyse médico-légale détaillée en cas de failles de sécurité majeures.

Les SIEM regroupent les données de plusieurs systèmes, ils les analysent pour détecter les comportements anormaux et les éventuelles cyberattaques ils collectent les événements et les alertes ils offrent une vue globale de l'environnement de sécurité dans l'organisation.

Le SIEM est important parce qu'il facilite la gestion de la sécurité par les entreprises en filtrant des quantités massives de données de sécurité et en hiérarchisant les alertes de sécurité générées par le logiciel.

Le logiciel SIEM permet aux organisations de détecter des incidents qui, autrement, pourraient passer inaperçus.

Le logiciel analyse les entrées du journal pour identifier les signes d'activité malveillante. En outre, comme le système rassemble des événements provenant de différentes sources sur le réseau, il peut recréer la chronologie d'une attaque, permettant ainsi à l'entreprise de déterminer la nature de l'attaque et son impact sur l'activité.

Un système SIEM peut également aider une organisation à répondre aux exigences de conformité en générant automatiquement des rapports qui incluent tous les événements de sécurité enregistrés parmi ses sources. Sans logiciel SIEM, l'entreprise devrait rassembler les données des journaux et compiler les rapports manuellement.

Un système SIEM améliore également la gestion des incidents en permettant à l'équipe de sécurité de découvrir la route qu'emprunte une attaque sur le réseau, identifier les sources qui ont été compromises et fournir les outils automatisés pour prévenir les attaques en cours.

Les limites d'un SIEM :

Malgré ses avantages, SIEM ne demeure pas sans limites dont les suivantes :

- Généralement, sa mise en œuvre prend beaucoup de temps, car elle nécessite une assistance pour garantir une intégration réussie avec les contrôles de sécurité d'une organisation et les nombreux hôtes de son infrastructure. Il faut généralement 90 jours ou plus pour installer SIEM avant qu'il ne commence à fonctionner.
- Le coût de mise en place est cher. L'investissement initial dans SIEM peut s'élever à des centaines de milliers de dollars. Et les coûts associés peuvent également s'additionner, y compris les coûts de personnel pour gérer et surveiller une implémentation SIEM, le support annuel et les logiciels ou agents pour collecter des données.
- L'analyse, la configuration et l'intégration de rapports nécessitent le talent d'experts. C'est pourquoi certains systèmes SIEM sont gérés directement au sein d'un centre d'opérations de sécurité (SOC), une unité centralisée dotée d'une équipe de sécurité de l'information qui traite les problèmes de sécurité d'une organisation.
- Les outils SIEM dépendent généralement de règles pour analyser toutes les données enregistrées. Le problème est que le réseau d'une entreprise génère un grand nombre d'alertes (généralement 10 000 par jour) qui peuvent être positives ou non. Par conséquent, il est difficile d'identifier les attaques potentielles en raison du nombre de journaux non pertinents.
- Un outil SIEM mal configuré peut manquer des événements de sécurité importants, ce qui rend la gestion des risques liés à l'information moins efficace.
- Ils peuvent être coûteux
- Gourmands en ressources

V.8 Les tendances futures du SIEM :

Actuellement, la SIEM ne fournit aux entreprises qu'une automatisation de base des flux de travail. Mais à mesure que les organisations continueront à se développer, SIEM devra offrir des capacités supplémentaires.

Par exemple, en raison de la commercialisation accrue de l'IA et de l'apprentissage machine, les outils SIEM devront offrir une orchestration plus rapide pour fournir aux différents départements d'une entreprise le même niveau de protection.

Les protocoles de sécurité et leur exécution seront plus rapides, plus efficaces et plus performants.

Meilleure collaboration avec les outils de détection et de réponse gérés (MDR).

Alors que les menaces de piratage et d'accès non autorisé continuent à augmenter, il est important que les organisations mettent en œuvre une approche à deux niveaux pour détecter et analyser les menaces de sécurité.

L'équipe informatique d'une entreprise peut mettre en œuvre le système SIEM en interne, tandis qu'un fournisseur de services gérés (MSP) peut mettre en œuvre l'outil MDR.

Amélioration de la gestion et de la surveillance des nuages.

Les fournisseurs de SIEM amélioreront les capacités de gestion et de surveillance du cloud afin de mieux répondre aux besoins de sécurité des organisations qui utilisent le cloud.

SIEM et SOAR vont évoluer vers un seul outil. Les produits SIEM traditionnels devraient bénéficier des avantages de SOAR, mais les fournisseurs de SOAR réagiront probablement en étendant les capacités de leurs produits.

V.9 Différence entre les SIEM et la gestion des logs :

Avant d'étudier les SIEM sur le marché, il est nécessaire de ressortir la différence entre SIEM et gestion des logs mais aussi d'expliquer leurs différences. La technologie SIEM assure la collecte, le regroupement, la normalisation et la conservation des logs adéquats, la collecte de données de contexte, l'analyse (y compris la corrélation et la hiérarchisation), la présentation (notamment la création de rapports et la visualisation), mais aussi les workflows et les contenus relatifs à la sécurité. Tous les cas pratiques d'utilisation de la technologie SIEM se concentrent sur la sécurité des informations, du réseau et des données mais aussi sur la conformité aux réglementations.

L'expression «gestion des logs» inclut la collecte exhaustive de logs, leur regroupement, la conservation des logs originaux (bruts, non modifiés), l'analyse du texte des logs, la présentation (principalement sous forme de recherches mais aussi de création de rapports) ainsi que les contenus et workflows connexes. Les cas pratiques d'utilisation des données de

consignation dans le cadre de la gestion des logs sont variés et couvrent tous les usages possibles des données de consignation dans tous les domaines possibles, bien au-delà du secteur informatique.

De ces deux définitions se détache une différence fondamentale : la technologie SIEM est orientée vers la sécurité (comme l'indique le premier mot de son acronyme anglais) et l'utilisation de diverses données informatiques à des fins de sécurité. En revanche, la gestion des logs est entièrement consacrée aux logs et au vaste éventail d'utilisations des données de consignation, au sein du secteur de la sécurité comme dans d'autres domaines.

VI. Solutions SIEM

VI-1-Comparaison entre les outils SIEM et la Gestion des logs

Dans le tableau ci-dessous, nous présentons les principales fonctions et illustrons les différences entre la gestion des logs et les outils SIEM.

Fonctionnalités	SIEM (Security Information and Event Management - gestion des informations et événements de sécurité)	Gestion des logs
Recueil de logs	Recueil des logs relatifs à la sécurité	Recueil de tous les logs, y compris les logs opérationnels et applicatifs personnalisés
Conservation des logs	Conservation limitée des données de consignation analysées et normalisées	Conservation plus longue des données de consignation brutes et analysées
Création de rapports	Création de rapports en temps réel à des fins de sécurité	Création de rapports génériques et historiques
Analyse	Corrélation, évaluation des menaces, hiérarchisation des événements	Analyse en texte intégral, étiquetage
Alertes et notifications	Création de rapports avancés à des fins de sécurité	Alerte simple pour tous les logs
Autres fonctionnalités	Gestion des incidents, autres analyses des données de sécurité	Évolutivité élevée en termes

Figure 2.a : Tableau comparatif SIEM et Gestion des logs

VI-2-Les principales solutions SIEM les plus utilisées du marché :

Une solution SIEM doit avant tout être évolutive car la réglementation et les sources de données peuvent évoluer. Elle doit être donc relativement simple à déployer et à maintenir, et doit pouvoir apporter le stockage à long terme des rapports associés.

Nous allons nous intéresser ici aux principales solutions SIEM du marché Open Source et Propriétaires.

a-Micro Focus ArcSight :

ArcSight Enterprise Security Manager possède des fonctionnalités de corrélation distribuée et de vue de cluster. Il est bon pour l'ingestion de sources car il prend en charge plus de 500 types d'appareils pour analyser les données. Il est disponible via l'Appliance, les logiciels, AWS et Microsoft Azure.

Fonctionnalités :

- Il fournit une corrélation distribuée en combinant le moteur de corrélation SIEM avec la technologie de cluster distribué,
- Il peut être intégré à diverses plates-formes d'apprentissage automatique et d'intelligence,
- Il utilise des agents ou des connecteurs. Il prend en charge plus de 300 connecteurs.

Micro Focus ArcSight est une solution évolutive pour répondre aux exigences de sécurité exigeantes. Il est bon pour bloquer les menaces et pour les performances (100 000 EPS).

b- Datadog :

Datadog Security Monitoring aide à la sécurité de la pile technologique grâce à la détection des menaces en temps réel. Datadog Security Monitoring unifie les développeurs, les opérations et les équipes de sécurité en une seule plateforme. Un tableau de bord unique affiche le contenu DevOps, les mesures commerciales et le contenu de sécurité.

Fonctionnalités :

- Avec plus de 400 intégrations soutenues par des fournisseurs, Datadog Security Monitoring permet de collecter des métriques, des journaux et des traces de l'ensemble de la pile ainsi que des outils de sécurité,
- Les règles de détection de Datadog offrent un moyen puissant de détecter les menaces de sécurité et les comportements suspects dans tous les journaux ingérés, en temps réel,

- Un éditeur de règles simple, pour répondre aux besoins spécifiques sans aucun langage de requête ;
- Datadog Security Monitoring brise les silos entre les développeurs, les équipes de sécurité et d'exploitation avec.

c-AT&T Cyber Security (AlienVault anciennement OSSIM : Open Source Security Information Management) :

AlienVault est une solution offrant une infrastructure pour le monitoring de la sécurité réseau. Ses objectifs sont :

- Fournir un cadre centralisé,
- Fournir une console d'organisation,
- Améliorer la détection et l'affichage des alarmes de sécurité.

Le but commun des trois principaux objectifs décrits ci-dessus est de permettre une réduction des faux positifs et des faux négatifs. Ce but est d'autant plus difficile à atteindre du fait qu'un volume considérable d'alarmes est présent. Il est donc nécessaire de relier ces différentes alarmes entre elles afin d'obtenir des informations pertinentes et d'éliminer les alarmes inutiles. Cet objectif peut être atteint à l'aide d'un procédé d'analyse des données nommée : corrélation. Le principe de fonctionnement d'OSSIM est formé de trois grandes catégories décomposées elles-mêmes en sous catégories :

La détection :

- IDS (pattern matching),
- Détection d'anomalie,
- Firewalls.

Corrélation :

- Evaluation de la dangerosité des alarmes,
- Evaluation du risque sur l'architecture à protéger les managements (configuration),
- Inventaire des ressources informatiques,
- Définition de la topologie,
- Définition des polices de sécurité,
- Définition des règles de corrélation, ´
- Liens avec différents outils d'audits Open Source.

Les différents procédés cités ci-dessus seront détaillées dans les sections suivantes.

La détection :

Celle-ci est évidemment effectuée à l'aide de sondes capables de traiter l'information en temps réel et d'émettre des alarmes lorsqu'une situation à risque est détectée.

Méthodes de détection :

Une sonde peut utiliser différentes approches afin de déterminer si un événement est à risque ou non. En effet, deux grands principes complémentaires (et non concurrent) sont présents dans la détection d'intrusions :

- Base sur des signatures
- Base sur la détection d'anomalie.

Détection par signatures:

La détection par signatures identifie des événements de sécurité qui tentent d'utiliser un système de façon non standard. Les représentations d'intrusions sont donc stockées et comparées à l'activité du système. Lorsqu'une intrusion connue est repérée lors de l'utilisation du système, une alarme est levée. La détection par signatures a des limites. Elle ne connaıt pas l'objectif de l'activité correspondant à une signature et va donc déclencher une alerte même si le trafic est normal. De plus, la détection par signature exige de connaitre préalablement l'attaque afin de générer la signature précise correspondante (fonctionnement par liste noire, exemple : antivirus). Ceci implique qu'une attaque encore inconnue ne pourra pas être détectée par signature.

Détection d'anomalie:

La détection d'anomalie identifie une activité suspecte en mesurant une norme sur un certain temps. Une alerte est ensuite générée lorsque le modèle s'éloigne de cette norme. Le principal avantage de la détection d'anomalie est qu'elle n'exige aucune connaissance préalable des attaques. Si le module de détection par anomalie détermine qu'une attaque diffère de façon significative de l'activité normale, il peut la détecter.

Comme la détection par recherche de pattern, la détection d'anomalie possède ses limites. Toute la difficulté réside dans la période de collecte des données correspondant à une activité désignée comme normale pour alimenter la base de données de référence.

L'analyse :

La méthode de traitement des données peut être décomposée en trois phases distinctes : ´

- Preprocessing : génération des alarmes et envoi de celles-ci.
- Rassemblement: toutes les alarmes précédemment envoyées (preprocessing) sont emmagasinées dans un serveur central.
- Post-processing : le traitement des données que l'on a emmagasinées. Les deux premières étapes (preprocessing et rassemblement) ne présentent rien de nouveau dans le cadre d'OSSIM. Celles-ci font principalement partie des méthodes de détection mentionnées ci-dessus (section). OSSIM offrira uniquement de nouvelles méthodes d'analyse et d'affichage des données récoltées. Il n'apportera en aucun cas de nouvelles solutions dans la détection et le rassemblement de données. Différentes méthodes d'analyse sont implémentées dans le cadre d'OSSIM :

– Définition de la priorité des alarmes,
– Evaluation des risques,
– Corrélation.

Définition de la priorité des alarmes:

Ce procédé permettra de déduire la priorité d'une alarme en fonction de son type, de la source de l'éventuelle attaque ainsi que de sa cible. Les exemples suivants illustrent facilement l'utilité de cette étape d'analyse:

– Une machine fonctionnant avec le système d'exploitation UNIX et hébergeant un serveur Web Apache est mentionné comme cible d'une attaque. Cette attaque a précédemment été détectée comme possible à l'aide d'un scanner de vulnérabilités. La priorité de l'alerte sera donc maximale puisque le serveur Web est vulnérable à l'attaque en question. Un procédé de cross-corrélation est donc appliqué à toutes les alertes. `

– Si la connexion à un serveur génère une alarme, la configuration de police de sécurité en fonction des utilisateurs permettra facilement la définition de la priorité de celle-ci.

En effet, différents cas sont envisageables:

– Priorité maximale de l'alarme si l'utilisateur (source de "l'attaque") est extérieur au réseau de l'entreprise et que la connexion a pour cible une base de données importantes.

- Basse priorité si l'utilisateur (source de "l'attaque") est interne au réseau de l'entreprise et que la connexion a pour cible une imprimante.
- Alarme discréditée si l'utilisateur (source de "l'attaque") fait partie de l'équipe développement et que la cible de la connexion est un serveur de test.

La définition de priorité fait partie d'une étape de mise en place de contextes des alarmes. Celle-ci est uniquement possible si l'environnement de l'entreprise est défini dans une base de données des connaissances. Cette base de données est définie par l'inventaire des biens informatiques ainsi que sur des polices d'accès à des serveurs et/ou services. La définition de ces différentes polices de sécurités implique la mise à disposition d'un outil de management permettant une configuration précise des polices de sécurité ainsi que des machines du réseau. Cette étape représente une part importante du filtrage des alarmes et nécessitera une constante mise à jour de la base de données des connaissances.

Evaluation des risques :

Le risque peut être défini comme étant la probabilité de menace de l'événement. En d'autres termes, cette étape tente de définir si la menace est réelle ou pas. L'importance à donner à un événement dépend principalement de trois facteurs :

- La valeur du bien attaque,
- La menace représentée par l'événement,
- La probabilité que l'événement apparaisse.

Les trois facteurs vus ci-dessus sont la base du calcul du risque intrinsèque.

Risque intrinsèque:

Ce risque peut être défini de la façon suivante : la mesure de l'impact potentiel de la menace sur le bien informatique, en fonction de la probabilité que cette menace apparaisse, tout en tenant compte de la fiabilité du capteur ayant émis l'alarme.

Le risque est traditionnellement représenté par le risque intrinsèque représentant le risque qu'une organisation court de par les biens qu'elle possède.

Risque immédiat :

Etant donné qu'OSSIM offre un traitement temps réel des alarmes, le calcul du risque immédiat peut être associe à la situation courante. Ce risque offre une vision de l'évaluation des dégâts qu'une alarme reçue pourrait engendrer. Cette évaluation prendra

aussi en compte la fiabilité du capteur ayant émis cette alarme. Le risque immédiat est donc calculé pour chaque alarme reçue et indique l'importance de l'alarme en termes de sécurité.

Corrélation :

Ce procédé permet notamment de retrouver certaines relations entre différentes alarmes indépendantes. Ceci permettra par exemple de découvrir des attaques noyées dans le flot des alarmes ou encore de discréditer des alarmes (découverte de faux positifs). La corrélation peut être simplement définie comme un procédé traitant des données (inputs) et retournant un résultat (outputs).

OSSIM utilise deux types d'inputs :

- Informations du moniteur (qui fournit normalement des indications à l'administrateur),
- Informations des détecteurs (qui fournissent normalement des alarmes).

Le résultat de ce traitement sera lui aussi d'un des deux genres cités ci-dessus (du moniteur ou des détecteurs). Les modèles de corrélations utilisés par OSSIM ont les objectifs suivants :

- Utilisation de méthodes par signatures, pour la détection d'événements connus et détectables
- Utilisation de méthodes sans signature, pour la détection d'évènements non connus
- Utilisation d'une machine d'états configurable par l'utilisateur, pour la description de signatures´ complexes
- Utilisation d'algorithmes évolués, pour l'affichage général de la sécurité

Méthodes de corrélation :

OSSIM met en œuvre plusieurs méthodes de corrélation complémentaires :

- Corrélation utilisant des séquences d'événements, basées sur les signatures connues et détectables,
- Corrélation utilisant des algorithmes heuristique, utilisée pour la détection d'attaques non connues,

- Cross-corrélation permettant la recherche de relations entre les scans Nessus effectués et des alertes détectées.
- Corrélation par heuristique OSSIM implémente un algorithme d'heuristique comme un accumulateur d'événements (CALM), offrant une indication de l'état général du réseau. L'objectif de ce traitement est d'obtenir en premier lieu, le risque immédiat puis le risque accumule.
- Le risque immédiat offre un haut niveau de monitoring temps réel. Celui-ci peut ´ être assimilé à un "thermomètre" des situations critiques, sans même connaitre les détails des caractéristiques du ´ problème.
- Le risque accumulé offre quant à lui un haut niveau de monitoring sur une certaine fenêtre temporelle (risque accumule et non plus temps réel).

Le second algorithme heuristique utilisé par OSSIM permet la prévision des statistiques réseaux (paquets émis et reçus) en fonction des valeurs précédentes (Holt-winter). Ceci permettra la détection automatique d'anomalies réseaux.

Par conséquence, la corrélation par heuristique offre :

- Une vue globale et rapide de la situation
- Une détection possible d'attaques, non relevées par les autres méthodes de corrélation (se basant ´ sur des signatures).

CALM (Compromise and Attack Level Monitor) est un algorithme utilisant les événements accumulés et fournissant une valeur indicative du niveau de sécurité global. L'accumulation d'événements est effectuée indépendamment pour tous les éléments du réseau. Celle-ci est simplement calculée par la somme de deux variables d'état représentant le risque immédiat de chaque événement.

Variable "C" ou niveau de fiabilité d'une machine ou d'un réseau, mesure la probabilité qu'une machine ou un réseau soit compromis (source d'une attaque effectuée par un ver ou troyen installé sur une machine à surveiller).

Variable "A" indique la probabilité que la machine ou le réseau à surveiller soit la cible d'attaques. La variable "A" représente la probabilité qu'une attaque a été lancée et qu'elle est réussie, alors que la variable "C" fournit l'évidence qu'il y a eu une attaque et qu'elle a réussi.

Chaque machine du réseau a donc une variable ”A” et ”C” associée, fluctuant de la manière suivante :

- Toute attaque possible d'une machine 1 (source) vers une machine 2 (cible) incrémentera le niveau ´”A” de la machine 2 et le niveau ”C” de la machine 1.
- Lorsqu'une réponse à une attaque est détectée (signifiant que l'attaque est réussie), le niveau ”C” des deux machines sera incrémenté.
- Lorsque l'événement est interne (source interne), seul le niveau ”C” de la machine source est incrémentée.

CALM fonctionne d'une manière temps réel (accumulation temps réel des évènements). Il peut aussi être intéressant d'observer ces statistiques dans une fenêtre temporelle (accumulation sur le temps). En effet, ceux-ci varieront grandement en fonction de la fenêtre utilisée puisqu'un fonctionnement temps réel (accumulation continue d'événements) aura pour effet de noyer certaines alarmes critiques dans la masse d'information. Nous pourrions assimiler le fonctionnement d'accumulation sur le temps à un zoom sur les statistiques temps réel. L'algorithme implémenté dans le moteur de corrélation temps réel d'OSSIM, permettant la découverte de comportements anormaux sans l'utilisation de seuils.

La détection anormale de comportements est décomposée en trois parties, chacune établie selon son prédécesseur :

- Un algorithme pour prévoir pas par pas les valeurs d'une série chronologique.
- Une mesure de déviation entre les valeurs prévues et les valeurs observées.
- Un mécanisme détectant lorsqu'une valeur est ”trop déviante” de la valeur prévue.
- L'algorithme de prévision Holt-Winter fait appel aux principes de lissage exponentiel (genre de moyenne mathématique permettant la prédiction de valeurs).

En effet, Holt-Winter fournit une méthode un peu plus complexe utilisant un triple lissage exponentiel.

Ce genre de corrélation fait appel à la détection par signatures. Ceci permettra à l'utilisateur (administrateur réseau en charge de la sécurité), de définir des règles de corrélation à l'aide des signatures disponibles.

Le monitoring : Le monitoring consiste en l'affichage des informations fournies. Les consoles de monitoring utilisent les différentes données produites par les procédés de corrélation pour la construction d'un affichage efficace et/ou résumé.

Moniteur d'utilisation, session et profile, OSSIM place une grande importance sur le monitoring détaillé de chaque machine et profile.

Il existe 3 différents types de monitoring dans OSSIM :

- Monitoring d'utilisation, offrant une vue générale d'une machine (comme SNMP le ferait)
- Monitoring de profile, offrant des informations spécifiques sur l'activité des utilisateurs, nous permettant d'établir un profil (pop, http, etc..) pour chaque utilisateur.
- Monitoring de session, offre un affichage temps réel des sessions en cours d'un utilisateur.
- Path monitor (moniteur de chemin) : Ce moniteur offre un affichage temps réel des chemins de l'information empruntés par des données émises par différentes machines sur le réseau. Il utilise les informations fournies par le moniteur de session (identifiant chaque lien présent sur le réseau) et par le moniteur de risque (fournissant le niveau de risque de chaque machine) afin de construire un affichage agréable.

Deux méthodes d'affichage et d'analyse sont disponibles :

- Hard Link analysis (Analyse des liens TCP) : Cette méthode affiche uniquement les sessions TCP courantes. Le but de celle-ci est de pouvoir observer la propagation d'une attaque ou d'un ver afin de déterminer un périmètre de sécurité.

- Soft Link analysis : Cette méthode d'analyse offre l'affichage de tous les liens perçus sur le réseau (UDP, TCP, ICMP inclus).
- Forensic console (Console légale) : Cette console offre l'accès à toutes les informations recueillies et stockées par le collecteur. Elle est donc un outil de recherche sur la base de données d'événements. Celle-ci permet une analyse à posteriori détaillée et approfondie des éléments réseaux. Cette console offre un aperçu de haut niveau de la sécurité. Cette console permet la définition de seuils générant des alarmes de haut niveau à destination de l'administrateur réseau en charge de la sécurité. L'affichage est simple est le plus concis possible et permet la visualisation des informations suivantes Monitoring constant du niveau de risque
- Monitoring constant du réseau (statistiques d'utilisation) :
- Monitoring des seuils définis :
- Monitoring des profils dépassant les seuils

d-SolarWinds :

SolarWinds fournit une solution de détection des menaces pour le réseau sur site via Log and Event Manager. Il dispose de fonctionnalités de surveillance des périphériques USB et de correction automatisée des menaces. Log and Event Manager propose de nouvelles fonctionnalités telles que le filtrage des journaux, la gestion des nœuds, le transfert de journaux, la console d'événements et l'augmentation de la limite de stockage.

Actuellement plus de 3 500 professionnels de la sécurité ayant des ressources restreintes font confiance à SolarWinds Log & Event Manager pour sa gestion des informations et événements de sécurité (SIEM) puissante, économique et efficace. C'est solution SIEM tout-en-un qui combine la gestion des journaux, la corrélation, la création de rapports, la surveillance de l'intégrité des fichiers, la surveillance des activités des utilisateurs, la détection et la prévention des problèmes liés aux ports USB, les renseignements sur les menaces et la réponse active dans une Appliance virtuelle facile à déployer, à gérer et à utiliser. C'est une solution SIEM conçue pour fournir les fonctionnalités dont les entreprises ont besoins sans la complexité et le coût des autres solutions SIEM d'entreprise.

Fonctionnalités principales:

Collecte aisée et évolutive des journaux des périphériques réseau, des ordinateurs et du Cloud :

Log & Event Manager collecte et répertorie les données des journaux et des événements en temps réel, quelle que soit l'origine des données générées au sein de l'infrastructure informatique.

Corrélation d'événements en mémoire et en temps réel :

En traitant les données des journaux avant qu'elles ne soient inscrites dans la base de données, Log & Event Manager offre une véritable corrélation des journaux et des événements en temps réel, permettant immédiatement d'intervenir et d'enquêter sur les violations de la sécurité et les autres problèmes critiques.

Flux d'informations sur les menaces :

Il s'appuie sur un flux intégré d'informations sur les adresses IP néfastes connues pour identifier les activités malveillantes. Ce flux est mis à jour régulièrement à partir d'une collection de sources de recherche et balise automatiquement les événements lorsqu'ils

entrent dans l'Appliance. À partir de là, nous pouvons exécuter rapidement des recherches ou des rapports pour voir les activités suspectes ou créer des règles pour l'exécution automatique d'actions.

Recherche informatique avancée pour l'analyse des événements :

La fonctionnalité de recherche informatique ad hoc de Log & Event Manager facilite la découverte des problèmes grâce à une interface de type glissé-déplacer qui suit instantanément les événements. Nous pouvons même enregistrer des recherches courantes pour une réutilisation ultérieure.

Compression et conservation des données des journaux :

Log & Event Manager stocke des téraoctets de données à un taux de compression élevé pour des rapports de conformité, des compilations et le déchargement, réduisant ainsi les exigences de stockage externe.

Surveillance incorporée de l'intégrité des fichiers en temps réel :

La surveillance incorporée de l'intégrité des fichiers offre une prise en charge élargie de la conformité et une intelligence de sécurité approfondie contre les menaces internes, les logiciels malveillants de type "jour zéro" et autres attaques avancées.

Selon les critiques, SolarWinds ne dispose pas d'une suite de sécurité complète mais offre de bonnes fonctionnalités et capacités de détection des menaces. Cela peut être une bonne solution pour les PME.

e-QRadar :

IBM QRadar SIEM est une plateforme de gestion de sécurité des réseaux offrant une prise en charge de la géolocalisation et de la conformité. QRadar SIEM combine à la fois la connaissance du réseau de flux, la corrélation des événements de sécurité et l'évaluation de la vulnérabilité des actifs.

IBM QRadar SIEM, permet de surveiller et afficher les événements de réseau en temps réel ou réaliser des recherches avancées. L'onglet Activité du journal affiche des données sur l'événement sous forme d'enregistrements provenant d'une source de journal, comme un pare-feu ou un routeur. L'onglet Activité du journal pour exécuter les tâches suivantes :

- Etudier les données d'événements,
- Etudier les journaux d'événements envoyés à QRadar SIEM en temps réel,
- Rechercher un événement,
- Surveiller les activités du journal à l'aide de graphiques en série temporelle configurables,
- Identifier les faux positifs pour régler QRadar SIEM. Pour plus d'informations, voir IBM QRadar - Guide d'utilisation.

IBM QRadar SIEM, permet aussi de mener des investigations sur les sessions de communication entre deux hôtes. Si l'option de capture de contenu est activée, l'onglet Activité réseau affiche des informations sur la façon dont le trafic réseau est transmis et quel est le contenu transmis. L'onglet Activité réseau vous permet d'effectuer les tâches suivantes :

- Etudier les flux envoyés à QRadar SIEM en temps réel,
- Rechercher des flux de réseau,
- Surveiller les activités réseau à l'aide de graphiques de série temporelle configurables.

Avec QRadar SIEM nous pouvons crée automatiquement des profils d'actifs en utilisant des données de vulnérabilité et des données de flux passives pour détecter les serveurs et hôtes de notre réseau. Les profils d'actifs fournissent des informations concernant chaque actif de notre réseau, y compris les services assurés. Les informations de profils d'actifs sont utilisées à des fins de corrélation et elles aident à réduire le nombre de faux positifs. Nous pouvons utiliser l'onglet Actifs pour exécuter les tâches suivantes :

- Rechercher des actifs,
- Afficher tous les actifs acquis,
- Afficher les informations d'identité pour les actifs acquis,
- Ajuster les vulnérabilités de faux positifs.

Dans IBM QRadar SIEM, nous pouvons mener des investigations sur les infractions afin de déterminer la cause première d'un problème dans le réseau.

Avec l'onglet Infractions pour afficher l'ensemble des infractions survenues sur le réseau et exécuter les tâches ci-après :

- Etudier les infractions, les adresses IP de source et de destination, les comportements de réseau ainsi que les anomalies de votre réseau.
- Mettre en corrélation les événements et les flux provenant de différents réseaux vers une même adresse IP de destination.

- Accéder aux différentes pages de l'onglet Infractions pour analyser les détails d'événement et de flux.
- Identifier quels sont les événements à l'origine d'une infraction.

Nous pouvons également créer des rapports personnalisés avec QRadar ou utiliser les rapports par défaut. QRadar SIEM fournit des modèles de rapports par défaut que nous pouvons personnaliser, rebaptiser et distribuer aux utilisateurs de QRadar SIEM. Les modèles de rapports sont regroupés par types de rapports, tels que les rapports de conformité, d'unité, de cadre ou de réseau.

L'onglet Rapports nous permet de réaliser les tâches suivantes :

- Créer, distribuer et gérer des rapports pour les données QRadar SIEM,
- Créer des rapports personnalisés à des fins d'exploitation et d'exécution,
- Combiner les informations de sécurité et de réseau dans un rapport unique,
- Utiliser ou éditer les modèles de rapport préinstallés,
- Baptiser nos rapports avec des logos personnalisés. Cette fonction est utile si nous souhaitons distribuer les rapports à différentes audiences.

QRadar SIEM accepte des informations dans différents formats et dans une large gamme de périphériques comme les événements de sécurité, le trafic réseau et les résultats d'analyse.

On distingue trois catégories de données collectées : la collecte des données d'événements, la collecte des données de flux et les informations d'évaluation de la vulnérabilité.

Collecte de données d'événement :

Les événements sont générés par des sources de journaux telles que les pare-feu, les routeurs, les serveurs et les systèmes de détection d'intrusion (IDS) ou les systèmes de prévention contre les intrusions (IPS). La plupart des sources de journaux envoie des informations à QRadar SIEM à l'aide du protocole Syslog.

QRadar SIEM prend également en charge les protocoles suivants :

- Simple Network Management Protocol (SNMP),
- JDBC (connectivité à la base de données Java),
- Security Device Event Exchange (SDEE).

Par défaut, QRadar SIEM détecte automatiquement les sources de journaux après un nombre spécifique de journaux identifiables reçus dans un intervalle de temps précis. Une fois les sources de journaux détectées, QRadar SIEM ajoute le module de prise en charge de

périphérique (DSM) approprié dans la fenêtre Sources de journal de l'onglet Admin. Même si la plupart des modules DSM comprend une fonction d'envoi du journal natif, certains requièrent une configuration supplémentaire, et/ou un agent, pour envoyer des journaux. La configuration varie d'un type de module DSM à un autre. Nous devons nous assurer que les modules DSM sont configurés pour envoyer des journaux dans un format pris en charge par QRadar SIEM. Certains types de sources de journaux, tels que les routeurs et les commutateurs, n'envoient pas assez de journaux pour que QRadar SIEM les détecte rapidement et les ajoute à la liste Source de journal. Nous pouvons ajouter manuellement ces sources de journaux.

En acceptant plusieurs formats de flux simultanément, QRadar SIEM peut détecter des menaces et des activités qui seraient sinon manquées en se basant strictement sur les événements d'informations.

Les Collecteurs QRadar QFlow offrent une détection complète des applications de trafic réseau quel que soit le port sur lequel l'application fonctionne. Par exemple, si le protocole Internet Relay Chat (IRC) communique sur le port 7500 (TCP), un QRadar QFlow Collector identifie le trafic en tant qu'IRC et fournit une capture des paquets du début de la conversation. NetFlow et J-Flow nous informent seulement qu'il y a du trafic sur le port 7500 (TCP), sans fournir de contexte quant au protocole utilisé. Les emplacements de ports de fonction miroir courants sont la mémoire système, DMZ, le serveur et les commutateurs d'application, NetFlow fournissant des informations supplémentaires provenant des routeurs et des commutateurs de limite.

Les Collecteurs QRadar QFlow sont activés par défaut et requièrent la connexion d'un port miroir, d'un port de réplication ou d'un tap réseau à une interface disponible du dispositif QRadar SIEM. L'analyse de flux commence automatiquement une fois le port miroir connecté à l'une des interfaces réseau du dispositif QRadar SIEM. Par défaut, QRadar SIEM surveille, sur l'interface de gestion, l'arrivée de trafic NetFlow sur le port 2055 (UDP). Nous pouvons affecter des ports NetFlow supplémentaires, si nécessaire.

Informations d'évaluation de la vulnérabilité :

QRadar SIEM peut importer des informations d'évaluation de la vulnérabilité (VA) de différents scanners tiers. Les informations VA permettent à QRadar Risk Manager d'identifier les hôtes actifs, les ports ouverts et les éventuelles vulnérabilités. QRadar Risk Manager utilise les informations VA pour classer l'ampleur des infractions sur notre réseau. En fonction

du type de scanner d'évaluation de la vulnérabilité, QRadar Risk Manager peut importer les résultats d'analyse depuis le serveur du scanner ou démarrer une analyse à distance.

f-Elasticsearch – Logstash – Kibana (ELK) :

ELK, maintenant appelé Elastic Stack, est une suite de logiciels édités par la société Elastic.

Elle permet en outre :

- la collecte des données au travers de Logstash (ou des Beats comme nous le verrons plus loin),
- le stockage des données dans le moteur d'indexation Elasticsearch,
- l'exploitation et l'analyse des données au travers de Kibana.

La mise en place d'un cluster est chose aisée ainsi que le déploiement dans le Cloud.

Plusieurs types d'architectures sont possibles, car nous verrons notamment que la collecte et le déversement des données peut se faire avec plusieurs types de sources ou puits de données.

Elasticsearch :

Elasticsearch est un moteur d'indexation basé sur Lucene. Il dispose d'une API REST et permet de faire du clustering très facilement, l'ajout d'un nouveau nœud étant très facile. En termes de requêtes, il est possible de faire des requêtes très simples comme des recherches sur des termes ou des chaînes de caractères, mais aussi de faire des agrégations ou encore de la percolation.

Les données sont stockées sous forme de documents dans des index Elasticsearch, un index étant lui-même découpé en types de documents. Attention, un type de document doit contenir des documents avec une structure commune.

En termes de découpage de données, nous pourrions avoir un index bibliothèque dans lequel nous retrouverons les types de document livre ou auteur… Concernant les logs, nous retrouvons souvent un découpage temporel ; par exemple, un index est créé pour chaque jour de données (ex : logstash-2020-08-20) dans lequel on retrouve des types de documents correspondant aux applications.

Les requêtes comme les réponses se font au travers de requêtes au format JSON avec la syntaxe Lucene ou en utilisant le Quercy DSL de l'outil (un futur langage spécifique à Kibana, Kuery, sera bientôt disponible).

Le logiciel est capable d'absorber des milliards de documents, tout est question de dimensionnement du cluster : pas de règle absolue, une étude devra être réalisée pour estimer les volumes et adapter le cluster en conséquence.

g-Logstash :

Historiquement dédié aux logs comme son nom l'indique, Logstash est un collecteur de données. Écrit en JRuby, son fonctionnement est décrit dans un fichier de configuration avec trois phases principales :

- Input : l'alimentation des données ; Logstash dispose de plugins permettant de se connecter à tout type de source de données (fichiers, base de données, broker…),
- Filter: c'est la partie la plus «intelligente» de Logstash, car c'est là où a lieu la transformation des données lues dans le cas où l'outil est utilisé pour écrire dans Elasticsearch, il sera donc nécessaire de transformer ces données en JSON. Très souvent, GROK [GROK] sera utilisé comme outil de parsing d'expressions régulières pour transformer facilement nos champs.
- Output : le déversement des données ; une fois lues et transformées, il est nécessaire de déverser les données. Ici encore, Logstash dispose de pléthore de plugins ; le plugin Elasticsearch est natif et il suffit de quelques lignes pour le configurer.

Kibana :

Kibana est l'interface web qui permet d'analyser toutes les données. L'outil s'occupe principalement de générer les requêtes Elasticsearch qui vont bien selon les demandes de l'utilisateur réalisées dans l'interface graphique.

Par défaut pour l'analyse, nous pouvons utiliser les fonctionnalités suivantes :

- Discover : requête permettant de trier/filtrer les données sur un index donné,
- Visualize : graphique permettant de montrer les données sous forme agrégée (camembert, cartographie, histogramme…),
- Dashboard : c'est la partie «métier» souvent présentée à nos clients : ce n'est ni plus ni moins qu'une page qui agrège des résultats de Discover ou encore des graphes réalisés dans la partie Visualize de l'application,

- DevTools : anciennement Sense dans les précédentes versions de Kibana ; il permet de requêter directement Elasticsearch (avec coloration syntaxique et indentation des requêtes).

Depuis les dernières versions, Kibana tend à être de plus en plus complet puisque nous avons à présent par défaut la présence du Timelion, un outil permettant de comparer facilement des données provenant d'index différents au même moment (par exemple, dessiner des «time séries» sur une fenêtre de temps donnée pour expliquer une consommation CPU trop élevée).

Divers : beats, X-Pack et licences :

Comme nous l'avons vu précédemment, Logstash permet de gérer l'alimentation et le déversement des données, mais il est également possible d'utiliser des beats. Ce sont des «shippers» écrits en Go qui permettent de répondre à des problématiques diverses et variées de manière efficace : collecte de métriques (TopBeat), de fichiers (FileBeat), de données réseau (PacketBeat), d'évènements Windows (Winlogbeat) et bien d'autres... [BEATS].

Logstash dispose évidemment d'un connecteur par défaut sur les beats pour s'alimenter de leurs données (transmises via TCP).

Il est maintenant temps de parler un petit peu de licence ; jusqu'à maintenant, tout ce que nous avons vu ne nécessitait pas d'achat particulier.

Cependant, nous n'aurons pas les fonctionnalités suivantes :

- **Sécurité** : assurée par Security (anciennement Shield), cela permet de filtrer les accès à Kibana, à Elasticsearch et de mettre en place des règles très fines (par exemple un utilisateur ne pourra pas voir un ensemble de champs d'un type de document particulier) ainsi que la communication HTTPS,
- **Alerting** : assuré par Watcher, cela permet notamment de pouvoir lancer des alertes en cas d'évènements particuliers que nous pouvons configurer sous forme de règles à appliquer sur nos données,
- **Monitoring** : anciennement Marvel, cela permet d'avoir une vue exhaustive sur l'état du cluster et des nœuds Elasticsearch et de pouvoir anticiper les éventuels besoins de capacité (ajout de nœuds),
- **Reporting** : générer et partager des rapports facilement, planifier l'envoi automatique de documents tout en gardant des historiques complets et accessibles,

- **Graph** : analyser les liens entre nos données, cette fonctionnalité est très utile pour la détection de fraudes ou pour permettre de mettre en place des recommandations,
- **Machine Learning** : grande nouveauté encore en bêta, le Machine Learning va nous permettre d'analyser les données et de détecter des comportements anormaux et surtout d'en extraire les parties prenantes.

Bien qu'il soit possible de développer des outils maison (authentification avec Kibana en utilisant un reverse proxy...) ou d'utiliser d'autres outils/plugins open source (SearchGuard pour assurer la sécurité), nous n'aurons pas la flexibilité et la robustesse des outils suscités.

Cet ensemble d'outils forme ce qu'on appelle le X-Pack, et l'utiliser est payant. Il nous donnera automatiquement accès au support quel que soit l'offre choisie (Gold, Platinum, Enterprise).

h-Splunk :

Splunk est développé par l'entreprise du même nom, fondée en 2003.

Splunk est née en 2007. A l'origine, le produit se présentait lui-même comme un moteur de recherche capable d'agréger et d'indexer les logs de tous les équipements IT de l'entreprise et de faciliter leur exploration par des recherches comme sur Google.

Aujourd'hui encore, Erik Swan, CTO et co-fondateur de Splunk, continue de décrire son produit comme «le Google des données machines». C'est une simplification qui ne traduit pas vraiment la diversité des usages d'un outil mais qui s'affirme désormais bien davantage comme une plateforme analytique spécialisée dans les données «machines», (celles générées par les ordinateurs, les appareils mobiles, les équipements réseau, les baies de stockage, les objets connectés dans toute leur diversité IoT ainsi que les apps exécutées par ces équipements) plutôt qu'un simple moteur de recherche.

Souvent perçu comme un outil de «Business Intelligence» pour l'IT, Splunk trouve ses usages bien au-delà de la DSI et du suivi de l'état du système d'information. Il apporte ses services aux équipes métiers qui ont besoin d'analyser les données de la consommation des clients sur leurs services Web, les impacts économiques de défaillances technique, les impacts des évolutions apportées à l'expérience utilisateur, etc.

Derrière la plateforme analytique spécialisée dans les données «machines» Splunk, se dissimulent plusieurs produits complémentaires et des solutions ciblées.

En voici un tour d'horizon exhaustif :

Splunk Entreprise est le produit phare et historique de l'entreprise. Facile à déployer, capable de monter en charge via des implémentations distribuées, Splunk Entreprise est au cœur de l'offre et s'affirme comme une véritable plateforme évolutive et programmable sur laquelle bâtir des solutions,

Splunk Cloud est la version SaaS de Splunk. Elle offre à toutes les entreprises un accès aux fonctionnalités de Splunk Entreprise à travers le Cloud, donc sans avoir à se soucier de la mise en œuvre d'une infrastructure pour l'héberger.

Servi par un SLA garantissant une disponibilité à 100% et une montée en charge qui peut dépasser les 10 To de données ingérées par jour, Splunk Cloud peut aussi être utilisé en mode hybride avec des Search Heads en local et des Indexers dans le Cloud, des Search Heads dans le Cloud et des Indexers en local ou un mixte complet local/cloud des composantes pour servir tous les scénarios de confidentialité et d'accessibilité mobile envisagés.

Splunk Light, petit dernier de la gamme, cherche à apporter une nouvelle option aux TPE/PME et à contrer la concurrence sur le marché des solutions d'analyses de données dédiées à l'IT.

Hunk est tout simplement une déclinaison de Splunk sur les infrastructures Big Data classiques.

Au-delà de ses plateformes analytiques, Splunk propose aussi des solutions bâties sur ces plateformes qui apportent des modules et tableaux de bord prédéfinis pour concrétiser instantanément certains scénarios typiques.

Splunk Enterprise Security (Splunk ES), première des solutions de haut niveau introduites par Splunk transforme la plateforme en l'un des plus puissants SIEM du marché.

De plus en plus adoptée comme base de SOC, la solution permet aux entreprises d'identifier les failles, d'évaluer des signaux faibles mis en évidence par différentes corrélations, de suivre l'évolution des cyber-attaques, etc.

Splunk ES offre une vision de bout en bout sur toutes les données de sécurité : utilisateurs, réseaux, terminaux, accès, données en transit. Il comporte des KPI/KSI sur les risques, des tableaux de bord et Workflows pour comprendre les impacts et contextes d'événements de sécurité et d'activités cybercriminelles.

Splunk Mint est né du rachat de Bug Sense. C'est un ensemble de technologies (SDKs et Splunk Apps) destinées à collecter, surveiller et analyser les données mobiles et plus particulièrement les données de fonctionnement (performance, crashs, usages, transactions, etc.) des apps mobiles. Une Splunk App assure l'ingestion des données et leur indexation dans l'architecture Splunk Enterprise ou Splunk Cloud.

Splunk IT Service Intelligence (Splunk ITSI) est un outil IT récemment introduit de monitoring et d'analyse des systèmes d'information.

Il permet une approche orientée «données» de la gestion d'une infrastructure (y compris hybride grâce à des ponts vers les principaux Clouds) qui offre à la fois des outils de surveillance de haut niveau et des outils de troubleshooting de bas niveau.

Splunk User Behaviour Analytics (Splunk UBA), fruit du rachat de Caspida en 2015, est une solution clé en main d'analyses comportementales pour repérer des cyber-attaques en cours et des menaces internes. S'appuyant sur les Data Sciences, le Machine Learning, et des corrélations avancées d'événements, Splunk UBA aide à repérer les comportements déviants, à détecter les APTs, à signaler les usages inhabituels de solutions SaaS, à découvrir les tentatives d'exfiltration de données, etc.

L'utilisation de Splunk est gratuite pour des quantités de données petites (jusqu'à 500 Mo par jour).

La version Enterprise est disponible gratuitement pendant 60 jours avec la même limite de 500 Mo par jour (plus de limite lorsqu'on paye la licence). Elle permet le clustering, le monitoring, l'accès à des applications "premium", le support…

Une version Cloud existe aussi, avec un prix dépendant de la quantité de données par jour.

Splunk fonctionne dans le navigateur assurant une compatibilité avec toutes les plateformes, le serveur pouvant soit être local, soit pouvant être mis sous forme de clusters pour effectuer des traitements distribués dans la version payante.

Index Volume	Perpetual License (per GB)	Annual Term License (per GB)	Volume Purchase Discount
1 GB/Day	$4,500	$1,800	0%
10 GB/Day	$2,500	$1,000	44%
50 GB/Day	$1,900	$760	58%
100 GB/Day	$1,500	$600	67%
>100 GB/Day	Contact sales for custom pricing with additional volume discounts		

Figure h.1 : Prix de Splunk en janvier 2015

Le principe de fonctionnement :

Les entrées du logiciel :

Splunk permet de charger des événements (par exemple des ventes) dans son moteur d'indexation :

- Soit en prenant un fichier plat en entrée (fichier Microsoft Excel, CSV, Open Document),
- Soit en le connectant à une base de données via un connecteur,
- Soit en récupérant les données entrant sur la connexion (via TCP/IP par exemple) pour en extraire des événements à ajouter au moteur.

Il existe tout un magasin d'application intégré dans le logiciel pour assurer toutes les connexions possibles avec les solutions du marché. Une large communauté s'est formée pour compléter en permanence ce magasin d'application.

Parmi les applications, en voici quelques exemples :

- On peut ajouter les fonds de carte Google Maps, pour rendre plus visuels encore les rapports, on peut localiser les faits et lire cette information facilement sur un plan,
- On peut facilement superviser des machines virtuelles VMware, ou un parc d'ordinateurs Windows en entreprise,
- On peut enrichir les données insérées grâce à certains services. Par exemple on peut ajouter des informations tirées des adresses IP.

Ces possibilités peuvent être étendues en développant avec le Splunk SDK qui permet de créer soi-même de nouvelles applications pour adapter exactement Splunk à ses propres besoins.

L'analyse des données :

Splunk analyse les données, et en particulier le timestamp qui est associé à chaque événement, et va ensuite les indexer dans son moteur, le Splunk Indexer.

On peut ensuite faire du reporting sur l'ensemble de ces événements.

Le reporting :

Le reporting est le fait de présenter sous forme de rapports des données concernant une entreprise, par exemple l'évolution des ventes, de manière générale ou selon un angle particulier, afin d'aider à la décision pour la stratégie de l'entreprise.

Grâce à un langage spécifique, le SPL, ou Search Processing Language, on peut, dans une barre de recherche, analyser les données de manière multi-dimensionnelle. On peut obtenir facilement et rapidement des tableaux ou des graphes.

On peut aussi créer des tableaux de bord, qui enregistrent des recherches et permettent ainsi de monitorer en direct les résultats.

Le Search Processing Language :

Le Search Processing Language, ou SPL, est un langage développé par Splunk, qui peut faire penser au SQL par ses possibilités, mais dont la syntaxe est totalement différente. Il permet de tirer au mieux parti des fonctions proposées par Splunk et de profiter pleinement de la puissance du moteur d'indexation.

Syntaxe de SPL:

Toute requête SPL commence par la source, c'est-à-dire l'ensemble des événements étudiés, suivie d'un pipe ou barre verticale : « | ». On peut enchaîner plusieurs requêtes simplement en ajoutant un pipe au début de chacune.

Les requêtes peuvent ensuite commencer par différents mots-clés :

- chart, pour obtenir un graphe,
- timechart, pour obtenir un graphe fonction du temps,
- top, pour obtenir les premiers résultats et beaucoup d'autres mots-clés…,
- On peut ajouter des paramètres à ces mots-clés, par exemple top limit=20 pour avoir non plus les 10 premiers résultats mais les 20 premiers.

Pour comparer une donnée par rapport à une autre, on va simplement utiliser le mot-clé by.

Le mot-clé count désigne le nombre d'événement, de la même façon que le fait COUNT (*) en SQL.

On peut créer des sous-requêtes dans les requêtes pour obtenir des traitements plus fins.

L'interface du logiciel :

Splunk est constitué d'un certain nombre d'écrans. Voici la description de ceux que nous allons utiliser.

Au lancement, l'écran principal donne accès aux applications installées dans la barre latérale gauche. Par défaut, la seule application installée est Search & Reporting.

A l'ajout de sources de données dans la partie centrale haute et à des rapports personnalisés (non encore ajoutés) dans la partie centrale inférieure.

L'application de recherche et de reporting dispose d'un champ de recherche qui permet d'entrer une requête en langage SPL. À droite de ce champ de recherche, on peut préciser la période de temps que l'on veut étudier (en temps réel : la dernière heure ou en temps relatif : un mois précis de l'année dernière).
La partie basse dispose de plusieurs onglets :

- Statistiques va afficher un tableau des résultats de la requête (sous forme paginée),
- Visualisation va afficher un graphe correspondant à ce tableau de résultat.

Splunk propose de choisir le type de graphe créé, comme le fait un tableur même s'il en conseille certains selon ce qui doit être affiché.

On peut créer des rapports, pour pouvoir enregistrer les recherches et voir dynamiquement les évolutions dans les données. Chaque rapport contient une et une seule recherche enregistrée.

On peut aussi créer des tableaux de bord, qui contiennent plusieurs rapports et ainsi avoir une vue d'ensemble des données très rapidement.

L'application Search & Reporting peut aussi être gérée de manière plus visuelle que le champ de recherche en utilisant le Pivot. Celui-ci permet de directement choisir ce que l'on souhaite voir en abscisse et en ordonnée de sorte que l'on peut avoir très rapidement un petit aperçu des données, sans même connaître le SPL.

SIEM vs Splunk:

Splunk 1, SIEM 0:

Les Splunk Apps permettent, en fin de compte, de dédier Splunk à un usage spécifique, au point que le produit est désormais considéré par Gartner comme l'un des principaux leaders du marché du SIEM. Il est même le seul leader à progresser dans le dernier quadrant.

Ce n'est pas illogique : Splunk transforme toute entrée Log en événement et l'art du SIEM consiste à rechercher des évènements. Les solutions Splunk ES et Splunk UBA viennent apporter l'intelligence nécessaire à cette base fondatrice.

Splunk : SIEM 3.0:

Splunk s'affirme comme l'un de ces SIEM «nouvelles générations» dans lesquels l'utilisation intensive de l'analytique permet de repérer les signaux faibles, de limiter automatiquement la remontée d'incidents non significatifs en termes de sécurité, de répondre en quasi temps-réel aux attaques, de pointer les comportements anormaux et de détecter les menaces induites par l'utilisation abusive de comptes à privilèges. L'intégration du Machine Learning au sein des modules Splunk ES et Splunk UBA a parallèlement déjà démontré son potentiel dans la détection de fraude et la détection d'activités malveillantes ou criminelles à travers le réseau de l'entreprise.

La visibilité en temps réel, sur toutes les activités des systèmes, réseaux, bases de données et applications, assure aux entreprises une connaissance instantanée du monde extérieur avec des informations sur les menaces, les réputations et les vulnérabilités. De même, il procure une visibilité sur les systèmes, les données, les risques et les activités se produisant en interne.

Les responsables informatiques bénéficient enfin d'un accès complet et corrélé au contenu et au contexte nécessaire pour prendre rapidement des décisions sur la base des risques, et ainsi investir au mieux leurs ressources dans un paysage de menaces dynamique. Cette caractéristique est indispensable dans le but d'étudier les attaques lentes et furtives, de rechercher des indicateurs de compromission ou de corriger les contrôles de conformité.

Avantages et inconvénients de Splunk :

Avantages :

- Les tableaux de bord de Splunk ne sont pas statiques : les données peuvent être monitorées en temps réel.

- Possibilité d'adapter à la taille de l'infrastructure : on peut l'utiliser pour une petite structure qui crée assez peu de données, comme on peut le mettre en clusters ou en version cloud si l'on a de gros lots de données à intégrer.
- Simplicité des fonctions de base. N'importe quelle personne peut créer des rapports, sans avoir besoin d'être analyste de données. Les requêtes sont en langage courant et l'interface est très dépouillée.
- Grande évolutivité grâce à la communauté autour qui fournit de nombreux connecteurs et extensions.
- Puissance du moteur d'indexation.

Inconvénients :

- Aucun dispositif d'ETL : pas de moyen simple de corriger les données, ou d'effectuer de jointures. Il faut donc avoir un intermédiaire qui fait ce travail pour fournir des événements en entrée,
- Nouveau langage à apprendre en parallèle de tous les autres,
- Prix au gigaoctet de données traité.

VI-3-Tableau comparatif des outils SIEM étudiés dans ce travail :

Le tableau décrit sur la figure 4.a ci-dessous fait état de la comparaison des SIEM étudiés dans notre travail sur la base des différentes fonctionnalités de base d'un SIEM.

Fonctionnalités		Plateformes					
		AlienVault	ArcSight	DataDog	Qradar	Solarwind	Splunk
Déploiement	Windows	✓	✓		✓	✓	✓
	Linux	✓			✓		✓
	Mac				✓		✓
	Cloud	✓	✓	✓	✓	✓	✓
	SaaS	✓	✓	✓	✓	✓	✓
	Web	✓	✓	✓	✓	✓	✓
Sécurité Réseau	Analyse de la vulnérabilité	✓					✓
	Contrôle d'accès						✓
	Intervention en cas de menace						✓
	Pare-feu						✓
	Rapports et analyses						✓
	Suivi des activités						✓
	Système de détection d'intrusion	✓				✓	✓
	VPN						✓
Outils SIEM	Analyse des comportements				✓		✓
	Analyse médico-légale					✓	✓
	Gestion des journaux	✓		✓		✓	✓
	Gestion des points de terminaison				✓		✓
	Monitoring réseau			✓	✓		✓
	Rapports de conformité	✓				✓	✓
	Suivi de l'activité des utilisateurs				✓	✓	✓
	Surveillance de l'intégrité des fichiers	✓				✓	✓
	Surveillance en temps réel	✓		✓	✓	✓	✓
	Sécurité des applications			✓			✓
	Threat intelligence	✓		✓	✓	✓	✓
	Analyse du réseau	✓	✓				✓

Figure 4.a : Tableau comparatif des SIEM étudiés

VI-4-Choix de la solution :

De l'étude comparative résumée sur la figure 4.a ci-dessus, il en ressort que Splunk est la plateforme SIEM la plus riche en termes de fonctionnalité. En plus, les avis des utilisateurs de ce dernier sont favorables et est classée au premier rang (pour la septième fois consécutive) des SIEM du moment (2020). Notre choix s'est donc porté sur Splunk pour le déploiement dans un environnement simulé.

Chapitre 4 : Déploiement de la solution et tests

VII. Déploiement de Splunk Entreprise Security :

Splunk Entreprise Security est un produit de Splunk Inc., une multinationale américaine (fondée en 2003 par Rob Das et Erik Swan, basée à San Francisco) spécialisée dans la production des logiciels de recherche, de surveillance et d'analyse de données massives générées par la machine via une interface web. Splunk Enterprise facilite la collecte, l'analyse et l'exploitation concrète de la valeur inexploitée des données machine générées par l'infrastructure technologique, les systèmes de sécurité et les applications métier, ce qui permet de disposer des renseignements nécessaires pour optimiser les performances opérationnelles ainsi que résultats commerciaux d'une entreprise.

VII-1-Architecture de Splunk ES :

Le diagramme de la figure à la page suivante illustre l'architecture Splunk dans son ensemble.

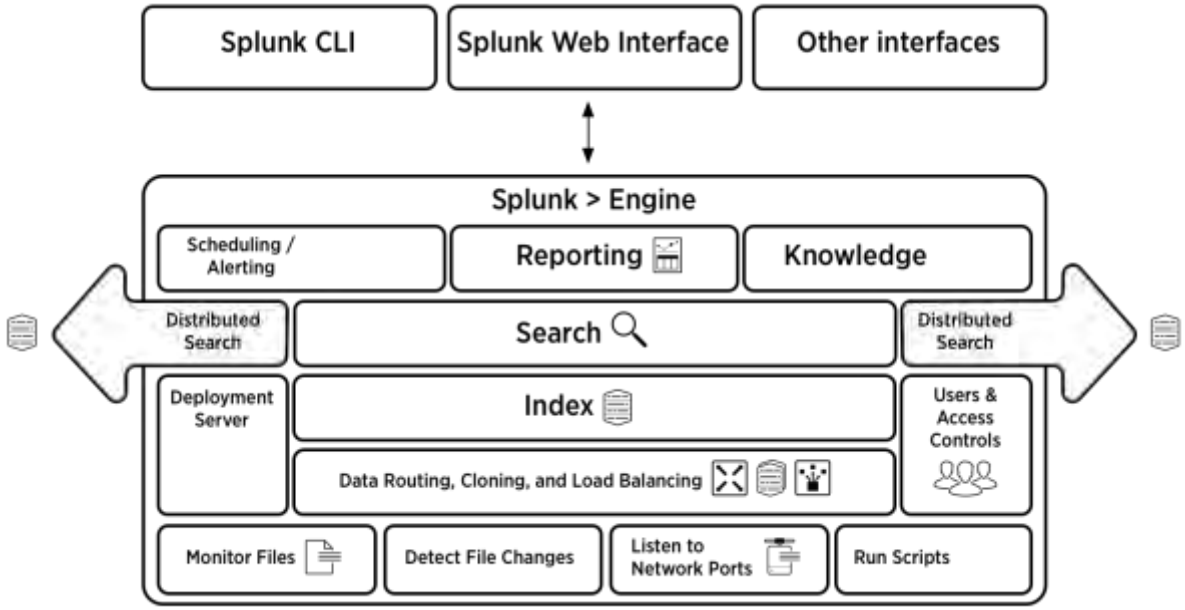


Figure 1.a : Architecture de Splunk ES [W6]

Du haut en bas, nous explorons dans les sous-sections ci-dessous les différents composants de cette architecture.

VII-2-Différentes étapes du pipeline de données :

Avant de parler du fonctionnement des différents composants Splunk, il est important de mentionner les différentes étapes du pipeline de données auquel chaque composant appartient. Il y a principalement trois étapes importantes dans Splunk à savoir :

- Etape d'entrée de données,
- Etape de stockage des données,
- Etape de recherche de données.

Etape d'entrée des données :

A cette étape, le logiciel Splunk consomme le flux de données brutes de sa source, le divise en blocs de 64 ko et annote chaque bloc avec des clés de métadonnées. Les clés de métadonnées incluent le nom d'hôte, la source et le type de source des données. Les clés peuvent également inclure des valeurs qui sont utilisées en interne, telles que le codage de caractères du flux de données et des valeurs qui contrôlent le traitement des données pendant la phase d'indexation, telles que l'index dans lequel les évènements doivent être stockés.

Etape de stockage des données :

Le stockage des données se compose de deux phases : l'analyse et l'indexation. En phase d'analyse, le logiciel Splunk examine, analyse et transforme les données pour extraire uniquement les informations pertinentes.

Ceci est également connu sous le nom de traitement des événements. C'est au cours de cette phase que le logiciel Splunk divise le flux de données en événements individuels. La phase d'analyse comporte de nombreuses sous-phases :

- Diviser le flux de données en lignes individuelles ;
- Identification, analyse et définition des horodatages ;
- Annotation d'événements individuels avec des métadonnées copiées à partir des clés à l'échelle de la source ;
- Transformer les données d'événement et les métadonnées selon les règles de transformation regex.

Dans la phase d'indexation, le logiciel Splunk écrit les événements analysés dans l'index sur le disque. Il écrit à la fois les données brutes compressées et le fichier d'index correspondant. L'avantage de l'indexation est que les données sont facilement accessibles pendant la recherche.

Etape de recherche de données :

Cette étape contrôle la manière dont l'utilisateur accède, affiche et utilise les données indexées. Dans le cadre de la fonction de recherche, le logiciel Splunk stocke des objets de

connaissance crées par l'utilisateur, tels que des rapports, des types d'événements, des tableaux de bord, des alertes et des extractions de champs. La fonction de recherche gère également le processus de recherche.

En résumé une solution de Splunk ES assure donc des fonctions de collecte d'événements en temps réel, de surveillance, de corrélation et d'analyse pour faciliter l'identification des menaces en temps réel résume dans le tableau ci-dessous :

Figure 2.a : Etapes de fonctionnement de Splunk ES [W13]

VII-.3-Composants Splunk :

Une étude détaillée de l'architecture Splunk met en évidence trois composants principaux dans Splunk à savoir :

- Splunk Forwarder : utilisé pour la transmission de données,
- Splunk Indexer : utilisé pour l'analyse et l'indexation des données,
- Search Head : est une interface graphique utilisée pour la recherche, l'analyse et la création de rapports.

Splunk Forwarder :

C'est un composant indépendant de l'instance Splunk principale, utilisé pour la collecte des journaux à partir d'une machine distante. Ces redirecteurs installés sur plusieurs machines, transmettront les données du journal à l'indexeur Splunk pour traitement et stockage.

Les Splunk Forwarder sont aussi utilisés à des fins d'analyse en temps réel des données.

Il existe plusieurs types de forwarders cités ci-après, mais le plus courant est l'Universal forwarder.

- **Universal Forwarder:** il contient uniquement les composants nécessaires pour transmettre les données brutes collectées à la source. Il s'agit d'un composant simple qui effectue un traitement minimal sur les flux de données entrants avant de les transmettre à un indexeur. Puisqu'il y a un traitement minimal sur les données avant leur transmission,

beaucoup de données inutiles sont également transmises à l'indexeur, ce qui entraine des surcouts de performances.

Pourquoi se donner la peine de transférer toutes les données vers les indexeurs puis de filtrer uniquement les données pertinentes ? Ne serait-il pas préférable d'envoyer uniquement les données pertinentes à l'indexeur et d'économiser de la bande passante, du temps et de l'argent ? Cela peut être résolu en utilisant des redirecteurs lourds ci-dessous.

- **Heavy Forwarder:** C'est une instance Splunk Enterprise complète qui peut indexer, rechercher et modifier les données ainsi que les transférer intelligemment vers l'indexeur, ce qui économise de la bande passante et de l'espace de stockage. Ainsi, lorsqu'un transitaire lourd analyse les données, l'indexeur n'a besoin que de gérer le segment d'indexation.

- **Light Forwarder:** Il est également une instance Splunk Enterprise complète, avec plus de fonctionnalités désactivées pour atteindre une empreinte de ressources aussi petite que possible. Le redirecteur léger est obsolète depuis la version 6.0 de Splunk Enterprise. Le transitaire universel remplace le transitaire léger à presque toutes les fins et représente le meilleur outil pour envoyer des données aux indexeurs.

Splunk Indexer :

L'indexeur est une instance Splunk Enterprise en charge de l'analyse et d`indexation des données reçues. Il transforme les données entrantes en événements et les stocke dans des index pour effectuer des opérations de recherche efficacement. A la réception des données d'un redirecteur universel, l'indexeur analyse d'abord les données, puis les indexent ensuite. L'analyse des données est effectuée pour éliminer les données indésirables. Mais, si nous recevons les données d'un transitaire lourd, l'indexeur indexera uniquement les données. Splunk traite les données entrantes pour permettre une recherche et une analyse rapides. Il améliore les données de différentes manières, comme :

- Séparation du flux de données en événements individuels et interrogeables,
- Créer ou identifier des horodatages,
- Extraction de champs tels que l'hôte, la source et le type de source,
- Exécution d'actions définies par l'utilisateur sur les données entrantes, telles que l'identification des champs personnalisées, le masquage des données sensibles, l'écriture de clés nouvelles ou modifiées.

L'application de règles de rupture pour les événements multi lignes, le filtrage des événements indésirables et le routage des événements vers des index ou des serveurs spécifiques.

Ce processus d'indexation est également appelé traitement d'événements.

Un autre avantage de Splunk Indexer est la réplication des données. On n'a pas à se soucier de la perte de données car Splunk conserve plusieurs copies des données indexées. Ce processus est appelé réplication d'index ou clustering d'indexeur. Ceci est réalisé à l'aide d'un cluster Indexer, qui est un groupe d'indexeurs configurés pour répliquer les données de chacun.

Splunk Search Head :

La tête de recherche fournit l'interface utilisateur que les utilisateurs peuvent utiliser pour interagir avec Splunk. Il permet aux utilisateurs de rechercher et d'interroger des données Splunk, et d'interfacer avec les indexeurs pour accéder aux données spécifiques qu'ils demandent.

Splunk fournit une architecture de recherche distribuée, qui vous permet d'évoluer pour gérer de gros volumes de données et de mieux gérer le contrôle d'accès et les données géo-dispersées. Dans un scénario de recherche distribuée, la tête de recherche envoie des demandes de recherche à un groupe d'indexeurs, également appelés pairs de recherche. Les indexeurs exécutent la recherche localement et renvoient les résultats à la tête de recherche, qui fusionne les résultats et les renvoie à l'utilisateur.

V II-8-Modes de déploiement et Configuration requise de Splunk ES

En fonction de nos besoins, nous pouvons déployer Splunk Enterprise en tant qu'instance unique ou créer des déploiements couvrant plusieurs instances, allant de quelques instances à des centaines, voire des milliers d'instances.

Déploiements à instance unique :

Dans les petits déploiements, une instance de Splunk Enterprise gère tous les aspects du traitement des données, de l'entrée à l'indexation jusqu'à la recherche. Un déploiement d'instance unique peut être utile à des fins de test et évaluation et peut répondre aux besoins d'environnements de la taille d'un service.

Déploiements distribués :

Pour prendre en charge des environnements plus vastes où les données proviennent de nombreuses machines, où on doit traiter de gros volumes de données ou lorsque de nombreux utilisateurs doivent rechercher les données, on peut faire évoluer le déploiement en distribuant des instances Splunk Enterprise sur plusieurs machines. C'est ce qu'on appelle un "déploiement distribué".

Dans un déploiement distribué typique, chaque instance de Splunk Enterprise effectue une tache spécialisée et réside sur l'un des trois niveaux de traitement correspondant aux principales fonctions de traitement :

- Niveau d'entrée de données,
- Indexer tiers,
- Niveau de gestion de la recherche.

Nous pouvons, par exemple, créer un déploiement avec de nombreuses instances qui résident sur le niveau d'entrée de données et n'ingérer que des données, plusieurs autres instances qui résident sur le niveau d'indexation et indexer les données, et une instance qui réside sur le niveau de gestion de la recherche et gère les recherches. Ces instances spécialisées sont appelées "composants".

Configuration matérielle requise :

Cela représente les spécifications d'instance de base minimales pour un déploiement Splunk Enterprise de niveau production, une instance unique.

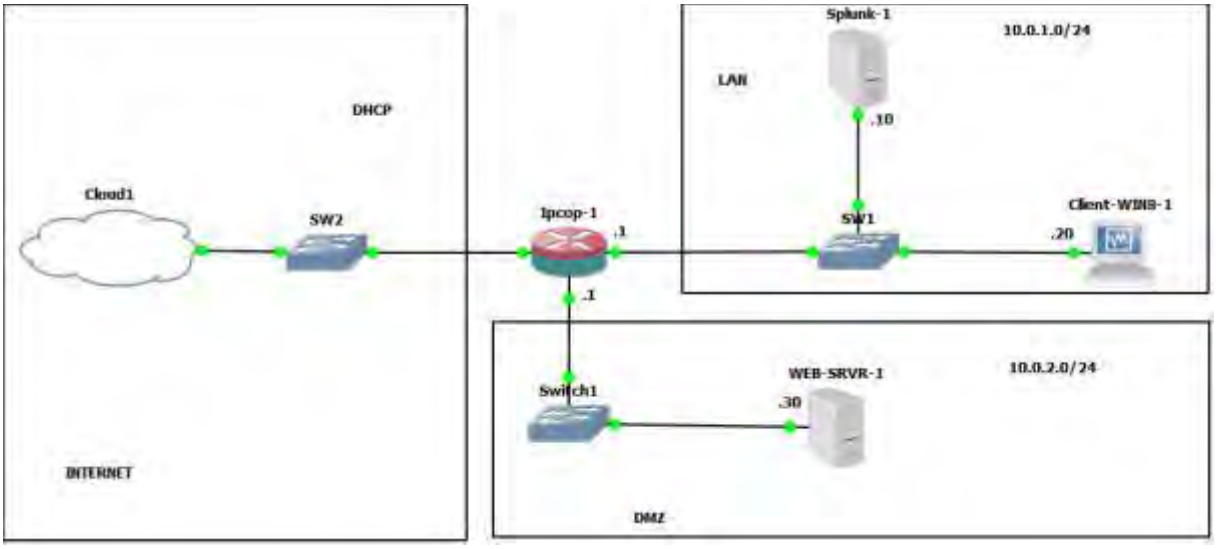
- Une architecture de puce x86 64 bits ;
- CPU 12 core de processeur physiques ou 24 vCPU à une vitesse de
- 2 GHz ou plus par cœur ;
- 12 Go de RAM ;
- Une carte réseau Ethernet 1 Go, une deuxième carte réseau en option pour un réseau de gestion ;
- Une distribution Linux ou Windows 64 bits.

Si on prévoit une installation Splunk Enterprise à une seule instance et que nous souhaitons une marge supplémentaire pour la concurrence de recherche ou plusieurs applications Splunk, envisageons d'utiliser les spécifications de milieu de gamme ou de haute performance de l'indexeur. Une fois que la capacité est dépassée d'un déploiement d'instance unique à répondre à notre charge de recherche et d'ingestion de données, si on examine les modèles de déploiement distribués pour une possible migration.

La section suivante est consacrée au de déploiement de l'application Splunk dans un environnement virtuel.

VII-9-Déploiement de Splunk ES

L'architecture ci-dessous correspond à un système basique adaptable à la plupart des systèmes en entreprise. Cette dernière a été mise en œuvre suivant les étapes décrites en annexe.



Fugure 5.a : architecture de déploiement.

L'architecture ci-dessus correspond à un système basique adaptable à la plupart des systèmes en entreprise. Cette dernière a été mise en laboratoire virtuelle à l'aide des outils Virtual Box et GNS3 suivant le plan d'adressage ci-après.

Poste	**SE et Version**	**Adresse IP**	**Masque**	**Passerelle**	**Interfaces**
Host	ElementaryOS Hera 5.1	172.16.31.1	255.255.255.0		vboxnet0
		DHCP	DHCP	DHCP	wl0
IPCFW	IPCop 2.1.8	172.16.31.2	255.255.255.0	172.16.31.1	rouge
		10.0.1.1	255.255.255.0		vert
		10.0.2.1	255.255.255.0		orange
WEB-SRVR	Ubuntu server 18.04	10.0.2.20	255.255.255.0	10.0.2.1	enp0s3
Splunk-Indexer	Ubuntu server 20.04	10.0.2.30	255.255.255.0	10.0.2.1	enp0s3

a-Préparation du déploiement :

Création et installation des machines virtuelles :

Les machines virtuelles sont créées suivant les étapes basiques de Virtual Box et d'installation.

Routage sur la machine hôte :

Après avoir installé et configuré nos machines virtuelles, nous devons les faire router sur notre hôte via l'IPCFW pour avoir accès à internet. Ceci grâce aux commandes suivantes :

Ajouter un résumé de route pour accéder aux réseaux 10.0.1.0/24 et 10.0.2.0/24 via le pare-feu avec la commande : **sudo route add -net 10.0.0.0/16 gw 172.16.31.2**

b- Installations des services sur les serveurs :

Installation des services apache, ssh et sur la machine WEB-SRVR

Nous avons installé le serveur web apache2 sur notre machine via la commande ***sudo apt-get install apache2*** pour pouvoir bénéficier des services web. Nous avons activé le module ssl au moyen des commandes suivantes :

sudo a2enmod ssl

sudo a2ensite default-ssl

sudo vim /etc/apache2/sites-available/000-default.conf

#Ajouter la ligne ci-dessous au VirtualHost

Redirect permanent / https://domaine

Le serveur ssh a été installé (et cela sur chaque serveur, ***apt-get install open-ssh***) pour pouvoir accéder au serveur depuis la machine hôte.

Passons ensuite au déploiement de Splunk à savoir l'installation de Splunk Entreprise Security et de Splunk Universal Forwader.

c- Installation de Splunk Entreprise Security sur la machine Splunk-Indexer :

L'installation s'est faite suivant les démarches ci-dessous.

- Téléchargement,
- Installation.

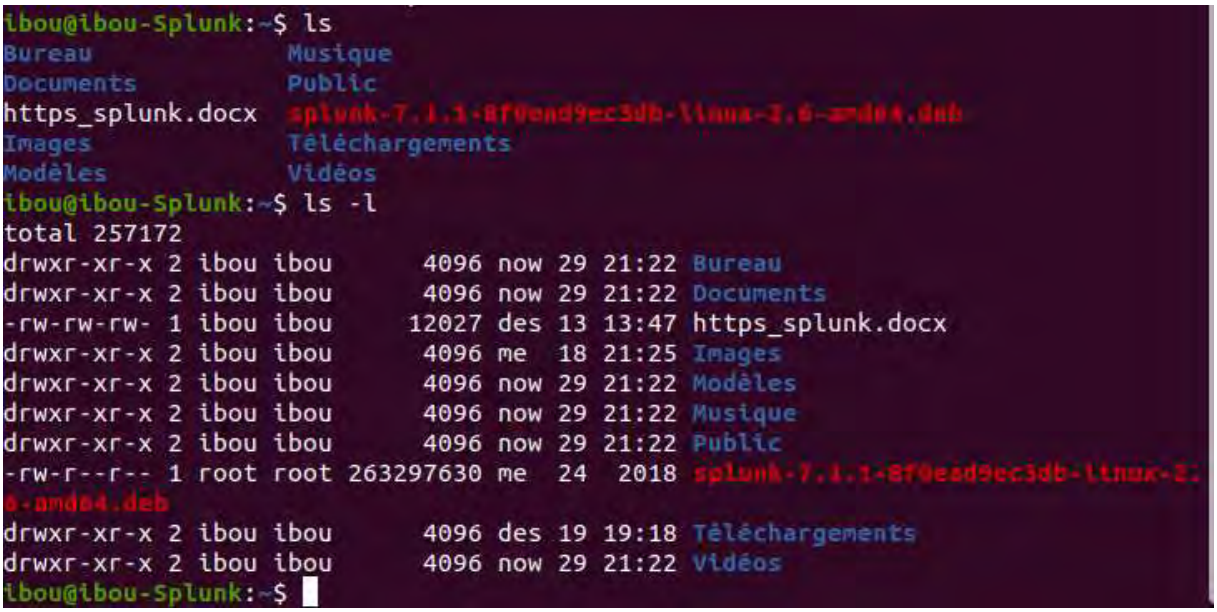


Figure 3.a : Téléchargement de Splunk.

- Démarrage de service Splunk, création d'utilisateur et mot de passe.

Figure 3.b : Installation de Splunk.

- Accès à Splunk depuis notre machine hôte via le navigateur.

d-Installation de Splunk Universal Forwarder sur la machine WEB-SRVR

Splunk Universal Forwarder doit être installé sur les machines dont les logs doivent converger sur le Splunk Indexer. Les étapes de cette installation sont ci-dessous.

- Téléchargement et Installations : Après avoir effectué le téléchargement au moyen de la commande wget,

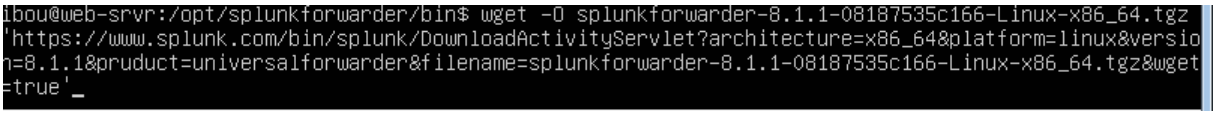


Figure 3.c : Téléchargement de Splunk Universal Forwarder.

Nous avons lancé notre installation comme ci-dessous.

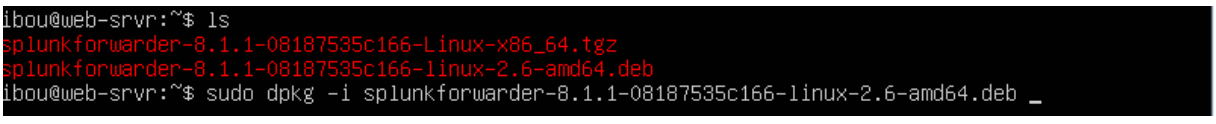


Figure 3.c : Installation de Splunk Universal Forwarder.

Figure 3.d : Interface web de Splunk.

- Lancement de forwarder : Nous lançons par la suite notre forwarder comme suit et activons son démarrage automatique au démarrage de notre machine virtuelle.

```
ibou@web-srvr:/opt/splunkforwarder/bin$ sudo ./splunk start --accept-license_
```

Figure 3.e : Activation de Forwarding des logs.

- Renseignement de l'indexer et du dossier des logs à envoyer : A ce niveau, nous avons tout d'abord renseigné l'indexer par son adresse et le port d'écoute (préalablement activé sur l'Indexer), pour ensuite renseigner le dossier des logs.

```
ibou@web-srvr:/opt/splunkforwarder/bin$ sudo ./splunk add forward-server 10.0.2.30:9997
```

Figure 3.f : Renseignement du dossier de logs

C'est ainsi qu'à pris fin la mise en place de notre laboratoire virtuelle devant nous servir pour la présentation du logiciel Splunk Entreprise Security. Passons à la section suivante pour la découverte de ce puissant outil SIEM.

e-Déploiement de Splunk Entreprise Security :

Dans cette sous-section, nous irons de l'interface d'accueil de notre application, à la création et démonstration des alertes, en passant par les recherches et la création des tableaux de bord.

Splunk ES Home :

L'interface d'accueil de Splunk Entreprise Security se présente comme sur l'image ci-dessous.

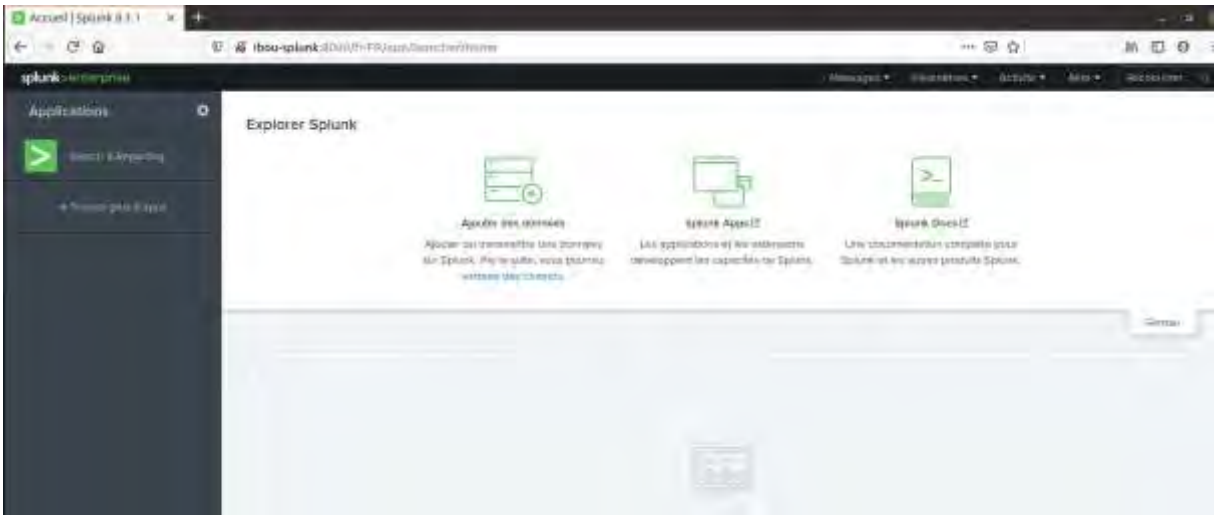


Figure 5.a : Accueil de Splunk ES

L'écran d'accueil comprend les éléments suivants :

- Barre de navigation (1) : Elle comprend des liens vers le profil utilisateur et les messages, paramètres et pages d'aide ;
- Menu Applications (2) : Il répertorie les applications installées sur notre instance Splunk que sous sommes autorisé à afficher. Dans notre exemple, seule l'application Search & Reporting est affichée ;
- Explorez Splunk Enterprise (3) : C 'est le panneau d'aide pour démarrer avec Splunk. Il permet de visiter le produit, d'ajouter des données à Splunk, d'accéder aux applications Splunk et à la documentation Splunk ;
- Tableau de bord d'accueil (4) : C'est dans ce panneau que nous pourrions ultérieurement ajouter notre propre tableau de bord personnalisé.

f-Ingestion de données :

Les données des serveurs sont envoyées instantanément à notre instance Splunk par le biais du Universal Forwarder grâce à la commande **$SPLUNK_HOME$/bin. /splunk add monitor « chemin ».** Il est aussi possible d'importer des données des fichiers externes (au format Excel par exemple) à des fins d'investigation, ce qui se fait à travers la fonction « **Ajouter des données** » disponible depuis l'Accueil Splunk et dans l'onglet **« Paramètres »**.

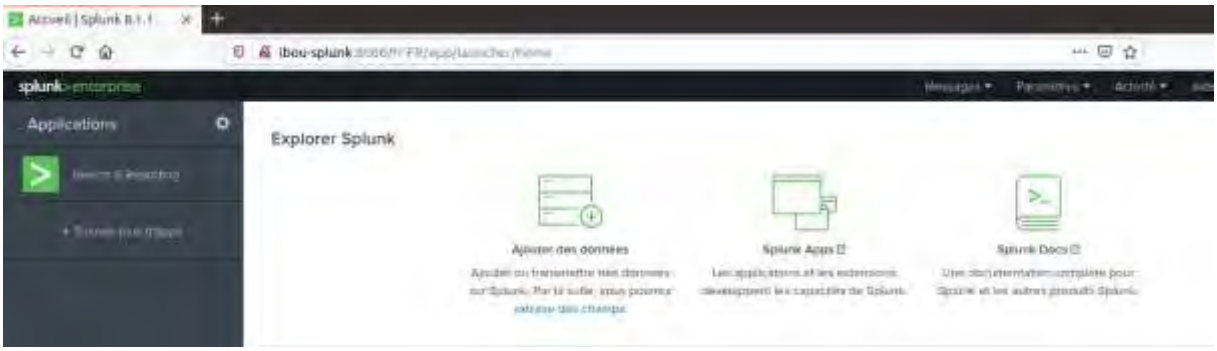


Figure 6.a : Fonction Ajouter des données

- En cliquant sur l'un de ces boutons, nous tombons sur les différentes options d'import des donnée, L'option « Envoyer » est celle de transfert des données interne.

Figure 6.b : Fonction Envoyer

- Après un clic sur Envoyer, l'écran nous montre les options de chargement de notre fichier : soit par sélection en suivant le chemin, soit par un simple glisser-déposer.

Figure 6.c : Selection de la source

Ensuite, nous passons à l'étape suivante en cliquant sur le bouton **« Suivant »**.

- Ce qui nous amène au paramétrage de la source comme indiquer sur la capture ci-dessous. Splunk détecte automatiquement notre type de fichier source et crée des champs appropriés à partir des données lues.

Figure 6.d : Paramétrage de la source

Si les paramètres sont corrects, nous passons à l'étape suivante, sinon, paramétrer puis cliquer sur « Suivant ».

- L'étape ci-après est le paramétrage d'entrée, nous avons laissé sur les options par défaut puis passés au suivant en cliquant sur « Résumé ».

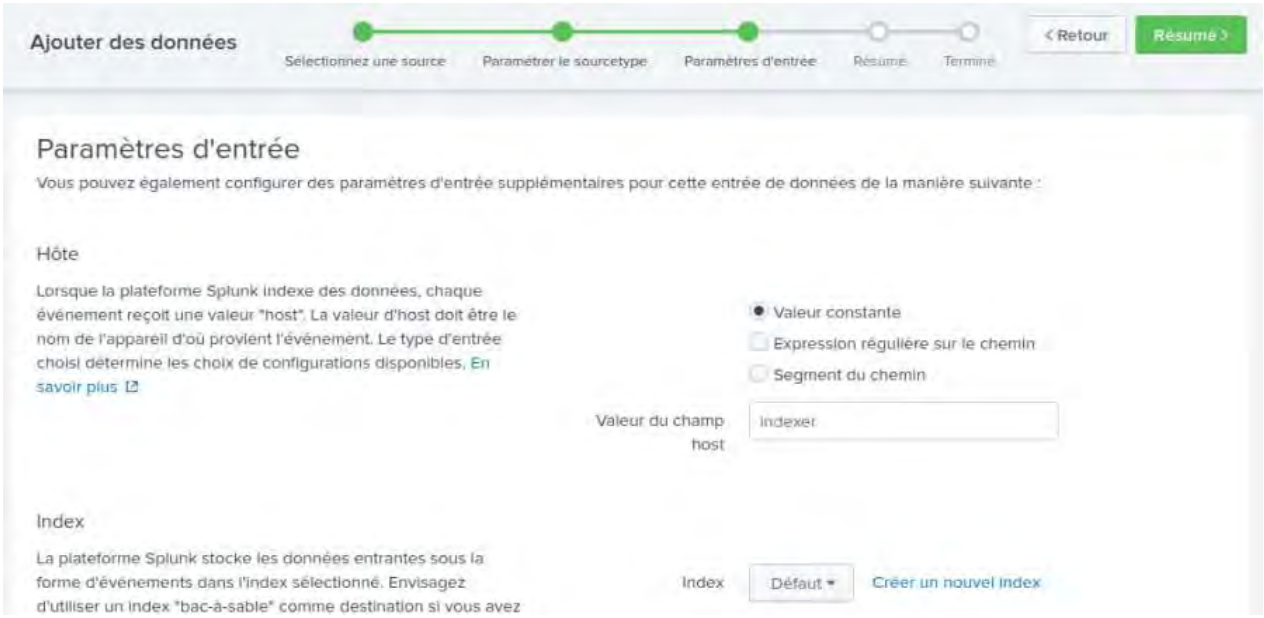


Figure 6.e : Paramètre d'entrée.

- L'étape ci-dessous nous donne un résumé global sur les propriétés de notre fichier.

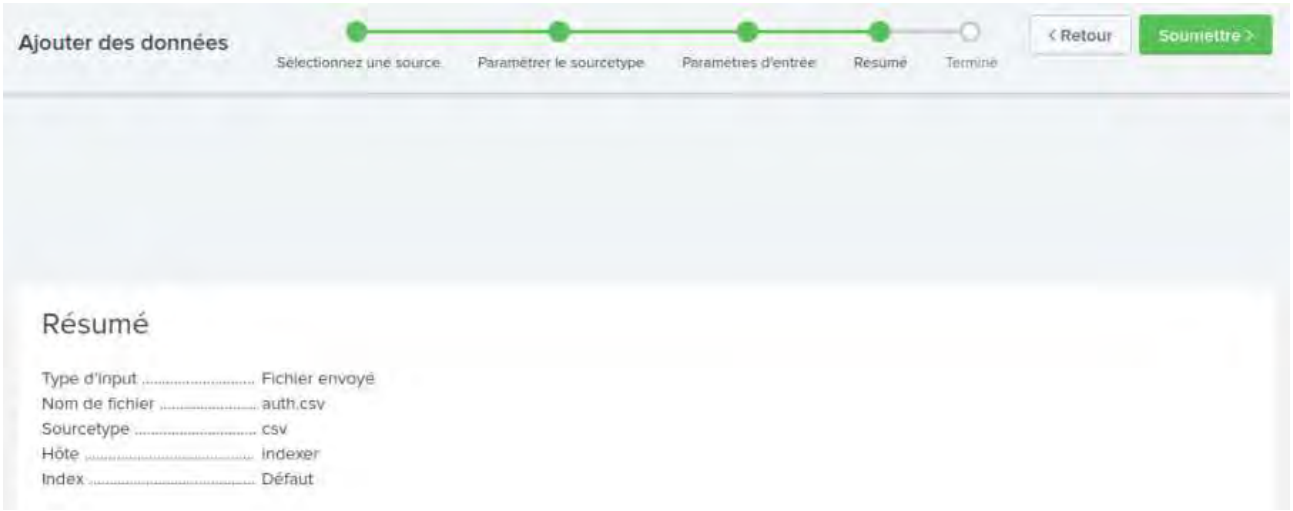


Figure 6.f : Résumé.

- En cliquant sur le bouton « Soumettre », nous validons notre importation et pouvons par la suite effectuer les opérations de recherche, d'extraction des champs etc. comme indiqué ci-dessous.

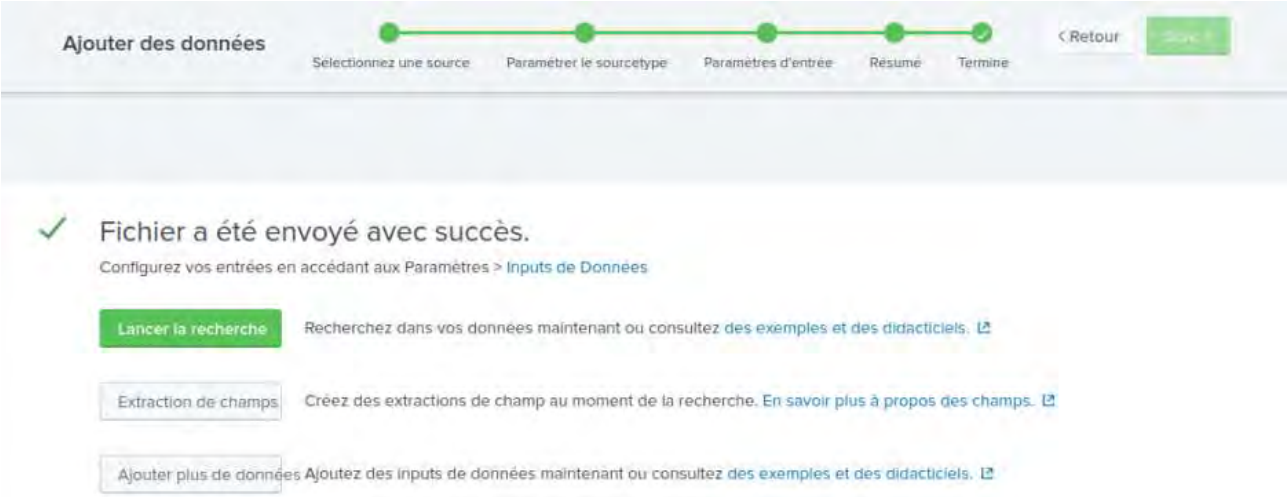


Figure 6.e : Envoi du fichier réussi

Après avoir fait l'injection de notre fichier, passons maintenant à la sous-section suivante consacrée à la recherche dans nos sources de données.

VII-10-La fonction Research de Splunk ES :

Splunk dispose d'une fonctionnalité de recherche robuste nous permettant de rechercher l'ensemble de données injectées dans Splunk. Cette fonctionnalité est accessible via l'application nommée **Search & Reporting** disponible dans l'accueil de notre application.

Figure 7.a : Accueil de recherche.

Vue Résumée de Recherche :

Lorsque nous cliquons sur l'application « Search & Reporting », nous nous retrouvons dans une résumée de recherche ci-dessous dans lequel nous pouvons commencer nos recherches sur les données téléchargées sur notre Indexeur dans la section précédente.

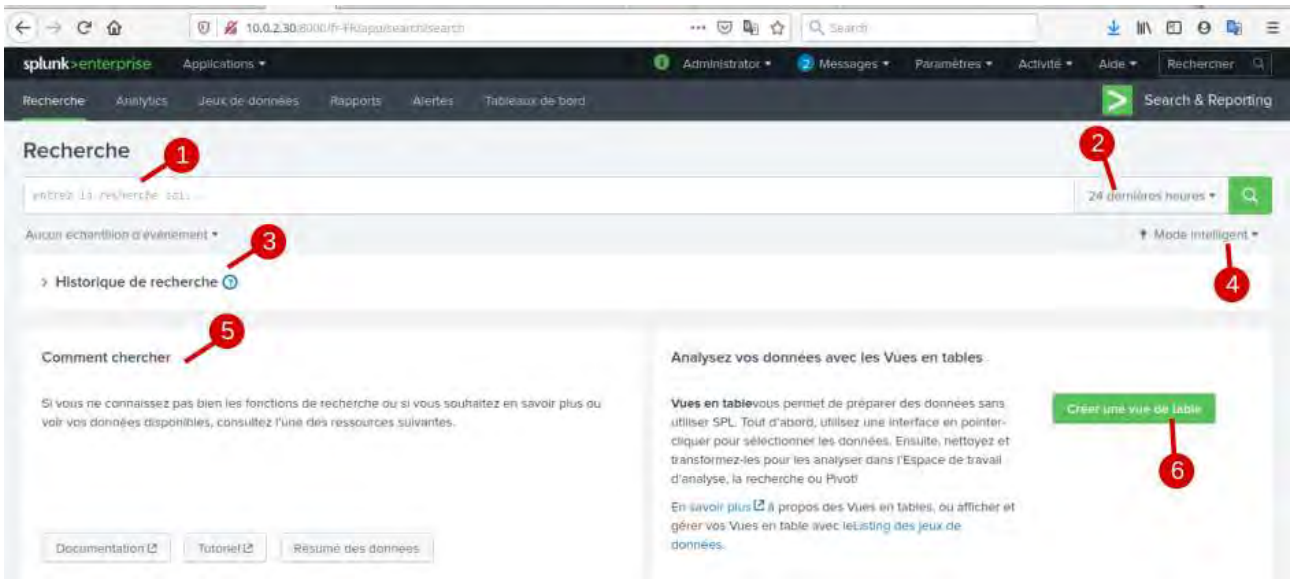


Figure 7.b : Nouvelle recherche.

La vue Résumé de recherche comprend :

- Barre de recherche (1) : Spécifie les critères de recherche ;
- Sélecteur de plage horaire (2) : Spécifie la période de recherche, par exemple les 30 dernières minutes ou hier. La valeur par défaut est « 24 dernières heures » ;
- Historique des recherches (3) : Affiche une liste des recherches effectuées précédemment ; L'historique de recherche apparaît après avoir exécuté votre première recherche ;
- Mode de recherche (4) : Permet de naviguer entre les modes Intelligent (par défaut), Rapide et Verbose ;
- Comment chercher (5) : Contient des liens vers la documentation et les tutoriels de recherche ;
- Créer une vue de table (6) : Permet de créer une vue personnalisée de table.

La vue Nouvelle recherche :

Elle s'ouvre après qu'on ait exécuté une recherche ou lorsqu'on clique sur l'onglet Recherche pour démarrer une nouvelle recherche et contient la plupart de boutons de la vue Résumée de recherche. Elle se présente comme ci-dessous.

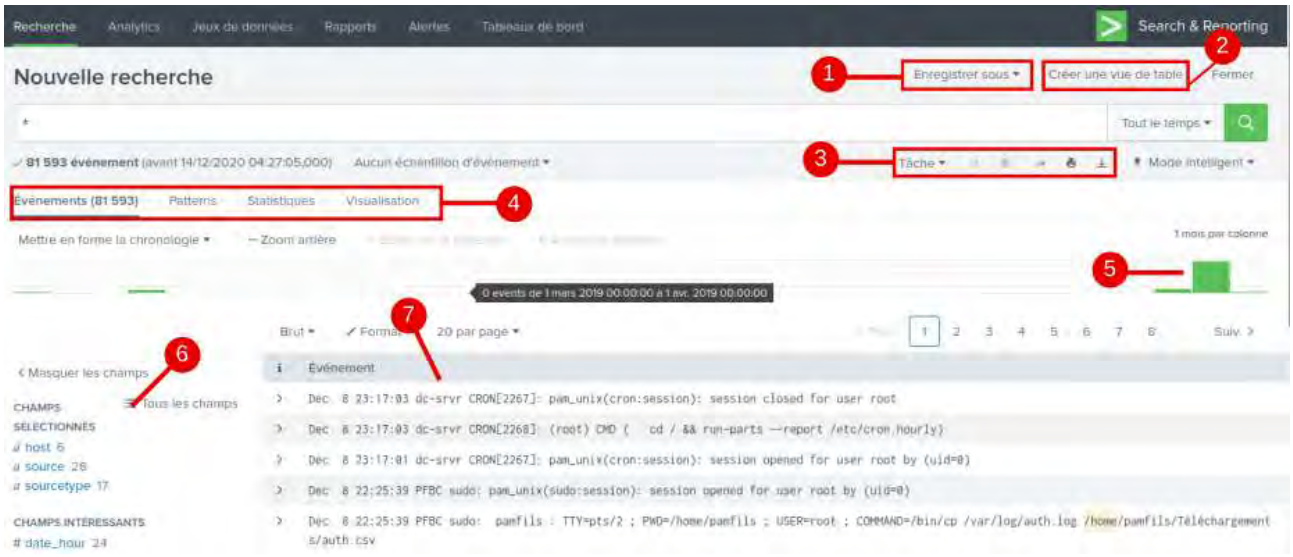


Figure 7.c : Termes de recherche.

Cette vue contient beaucoup plus d'éléments à savoir :

- Enregistrer sous (1) : Utilisé pour enregistrer les résultats de recherche sous forme de rapport, de tableau de bord, d'alerte ou de type d'événement ;
- Créer une vue de table (2) : Permet de créer une vue personnalisée de table ;
- Boutons d'action (3) : Permet d'effectuer des tâches, de partager, d'imprimer et d'exporter les résultats de recherche ;
- Onglets des résultats de recherche (4) : Ils sont déterminant dans l'affichage des résultats de recherche soit en Événements, soit en Patterns, Statistiques ou Visualisation ;
- Chronologie (5) : Une représentation visuelle du nombre d'événements qui se produisent à chaque instant. Les pics ou les creux dans la chronologie peuvent indiquer des pics d'activité ou des temps d'arrêt du serveur ;
- Barre latérale des champs (6) : Elle affiche une liste des champs découverts dans les événements. Les champs sont regroupés en « champs sélectionnés » et « champs intéressants » ;
- Visionneuse d'événements (7) : Affiche les événements correspondants à la recherche. Dans chaque événement, les termes de recherche correspondants sont mis en évidence. Cet affichage est modifiable en utilisant Liste, Format et Options par page.

Langage de recherche Splunk :

Une fois nos données indexées, nous pouvons commencer la recherche. Grâce à la recherche, Splunk permet à l'utilisateur de parcourir les données indexées pour trouver ce dont il a besoin pour répondre aux questions. Les recherches sont effectuées à l'aide d'un langage spécifique à Splunk : le Splunk Search Processing Language (SPL).

SPL est un langage contenant de nombreuses commandes, fonctions, arguments, etc., qui sont écrits pour obtenir les résultats souhaités à partir des ensembles de données.

Le SPL comprend les composants suivants :

- Termes de recherche : Ce sont les mots-clés ou expressions recherchés que nous mentionnons dans la barre de recherche pour obtenir des enregistrements spécifiques de l'ensemble de données répondant aux critères de recherche. Dans l'exemple ci-dessous, nous recherchons des enregistrements contenant deux termes en surbrillance.

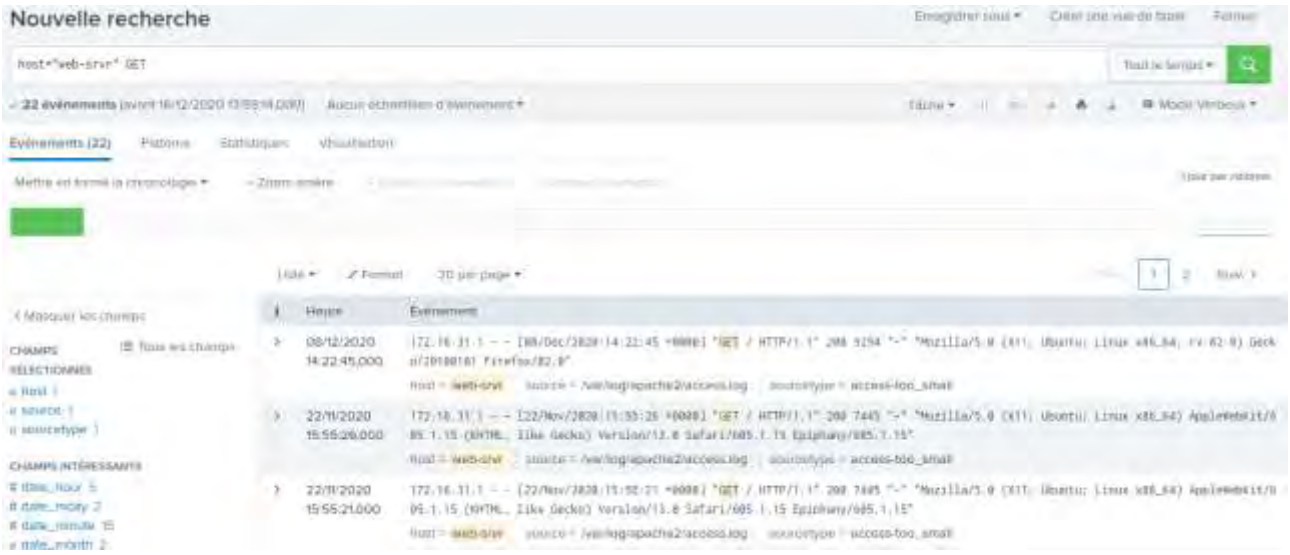


Figure 7.b : Nouvelle recherche.

- Commandes : Ce sont les actions à effectuer sur l'ensemble de résultats, comme supprimer les résultats ou les trier. Nombreuses commandes intégrées fournies par SPL peuvent être utilisées pour simplifier le processus d'analyse des données dans l'ensemble de résultats. Dans l'exemple ci-dessous, nous utilisons la commande **head** pour filtrer uniquement les 3 premiers résultats d'une opération de recherche.

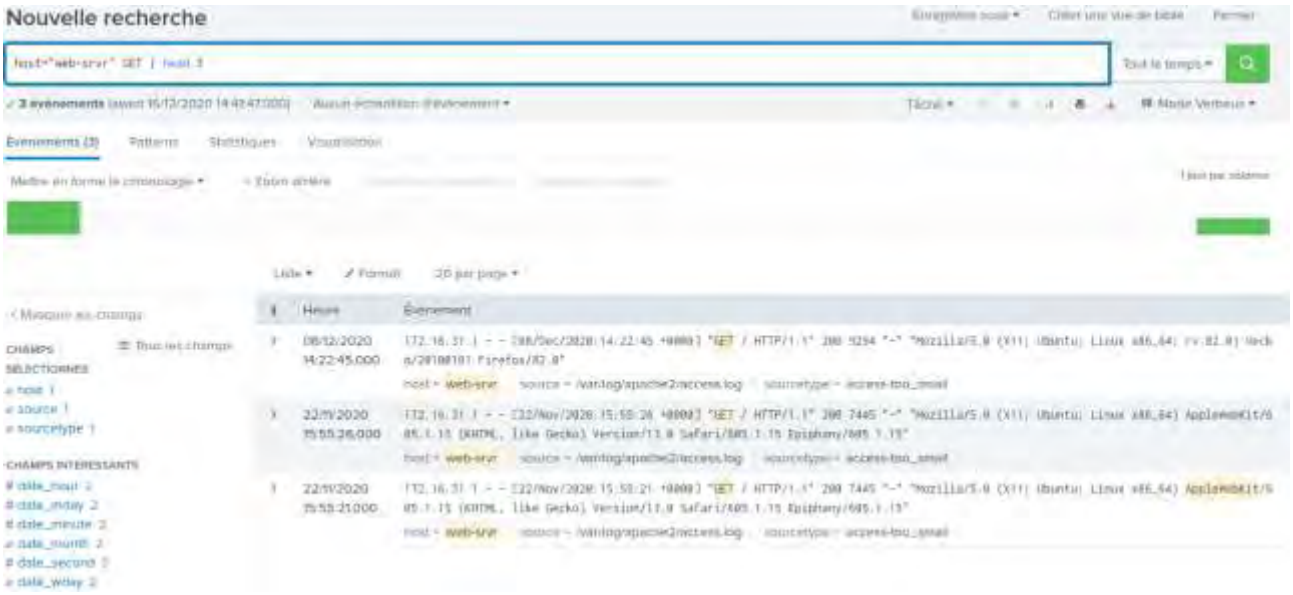


Figure 7.c : Commande head.

- Fonctions : Elles définissent quelles sont les opérations à appliquer sur les résultats. En plus des commandes, Splunk fournit également de nombreuses fonctions intégrées. Dans l'exemple ci-dessous, nous utilisons la fonction Stats avg() qui calcule la valeur moyenne du champ numérique pris en entrée.

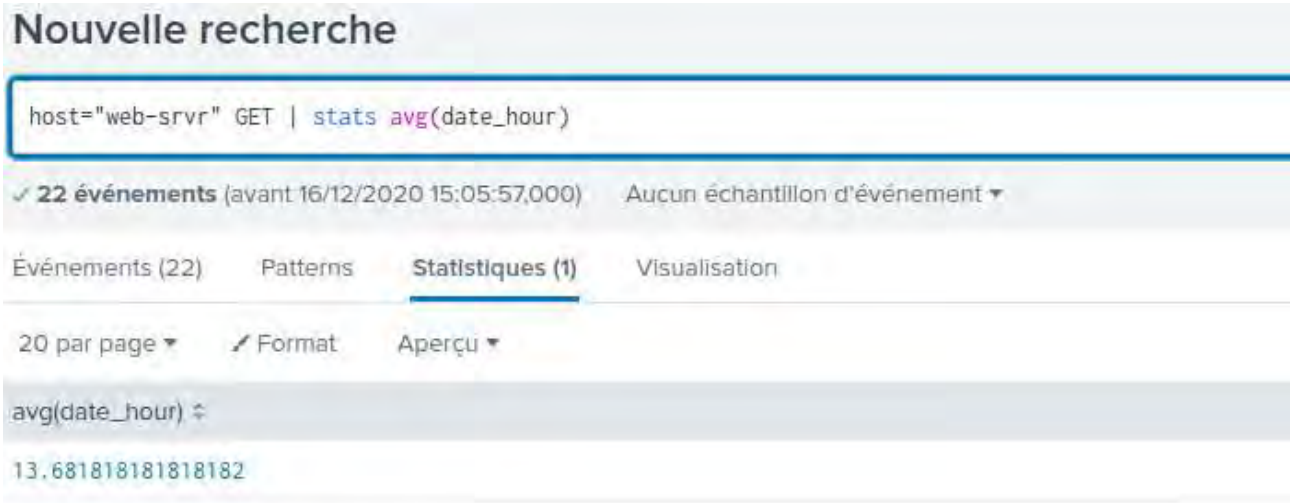


Figure 7.d : Fonction AVG.

- Clauses : définissent comment grouper ou renommer les champs dans le jeu de résultats. Lorsque nous voulons obtenir des résultats regroupés par champ spécifique ou que nous voulons renommer un champ en sortie, nous utilisons respectivement la clause **by** et la clause **as**. Dans l'exemple ci-dessous, nous obtenons le nombre des occurrences par hôtes. Le résultat montre le nom de chaque hôte ainsi que le nombre des lignes correspondantes.

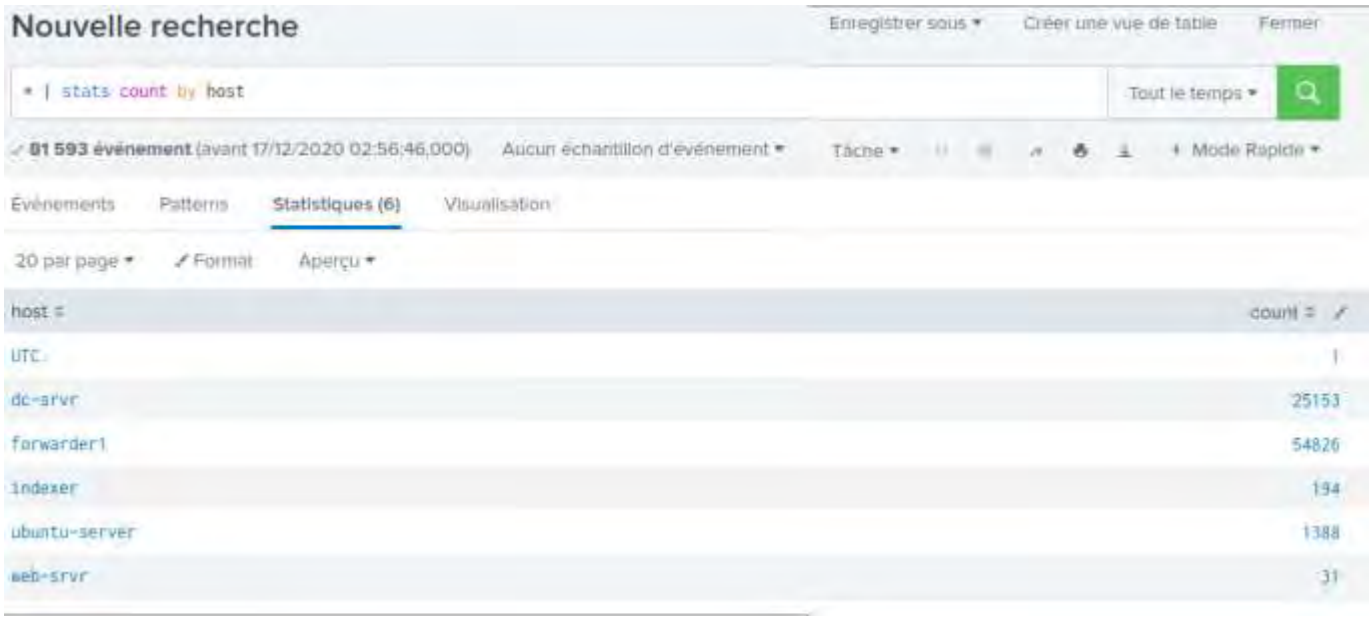


Figure 7.d : Fonction Count by.

Création des Rapports :

Les rapports Splunk sont des résultats enregistrés à partir d'une action de recherche qui peuvent afficher des statistiques et des visualisations d'événements. Ils peuvent être exécutés à tout moment et récupèrent de nouveaux résultats chaque fois qu'ils sont exécutés. Les rapports peuvent être partagés avec d'autres utilisateurs et peuvent être ajoutés aux tableaux de bord.

Dans cette sous-section, nous verrons comment créer et modifier un exemple de rapport.

L'option Enregistrer-Sous sur la figure ci-dessous est utilisée pour enregistrer les résultats d'une recherche en rapport.

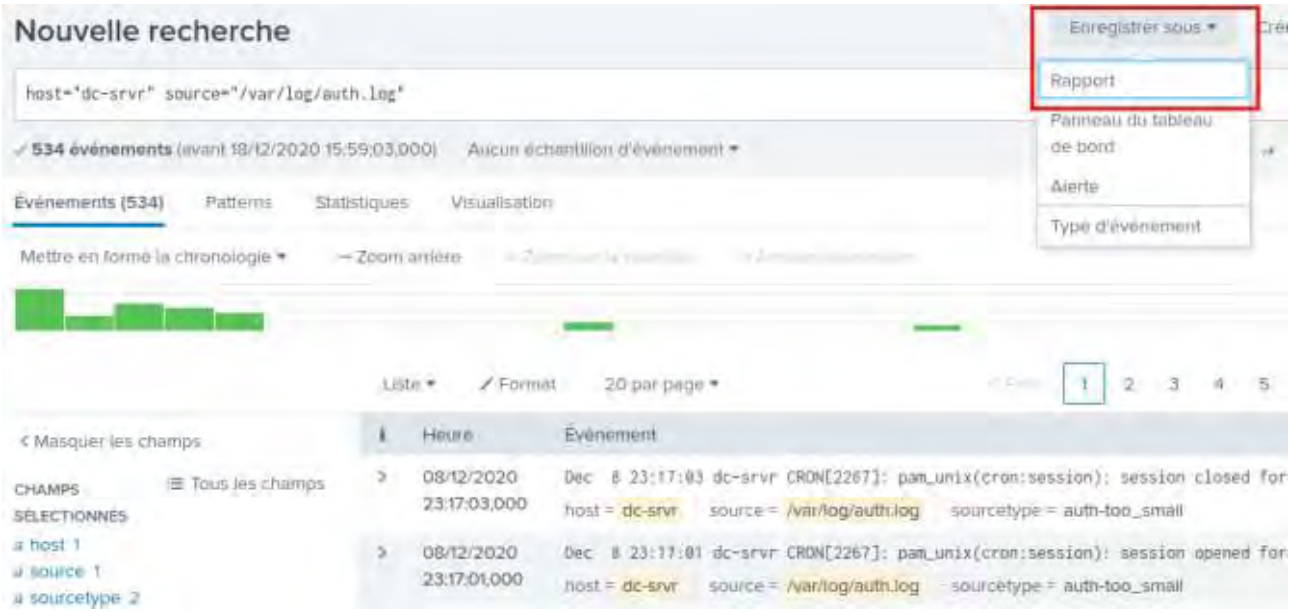


Figure 7.e : Crétion de rapport.

En cliquant sur l'option Rapports dans la liste déroulante, nous obtenons la fenêtre suivante qui demande des entrées supplémentaires comme le nom du rapport, la description et le choix du sélecteur de temps. Si nous choisissons le sélecteur de temps, il permet d'ajuster la plage de temps lorsque nous exécutons le rapport.

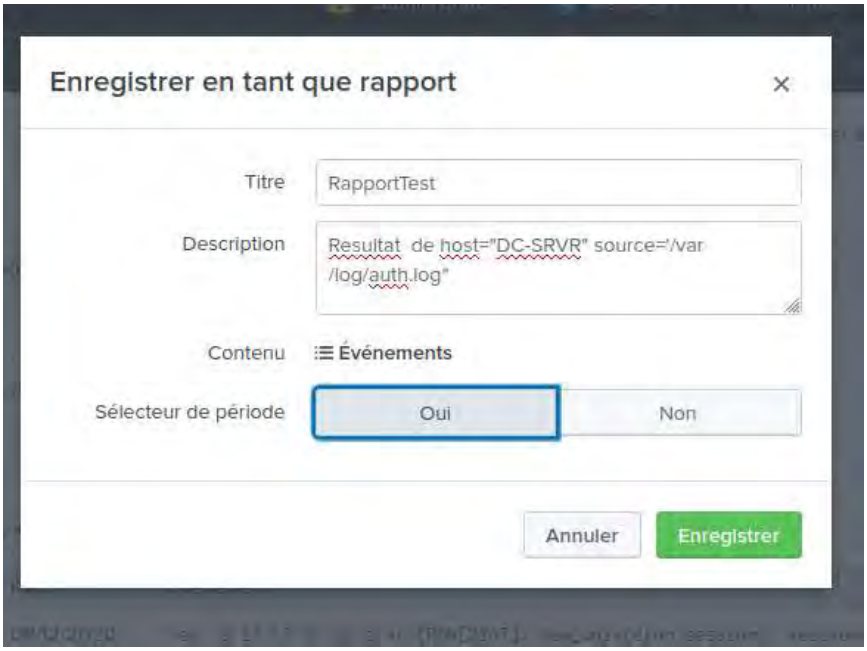


Figure 7.f : Paramétrage de rapport.

Après avoir cliqué sur Enregistrer pour créer le rapport à l'étape ci-dessus, nous obtenons l'écran suivant demandant de configurer le rapport comme indiqué ci-dessous. Ici, nous pouvons configurer les autorisations, planifier le rapport, etc. Nous avons également la possibilité de passer à l'étape suivante et d'ajouter le rapport à un tableau de bord.

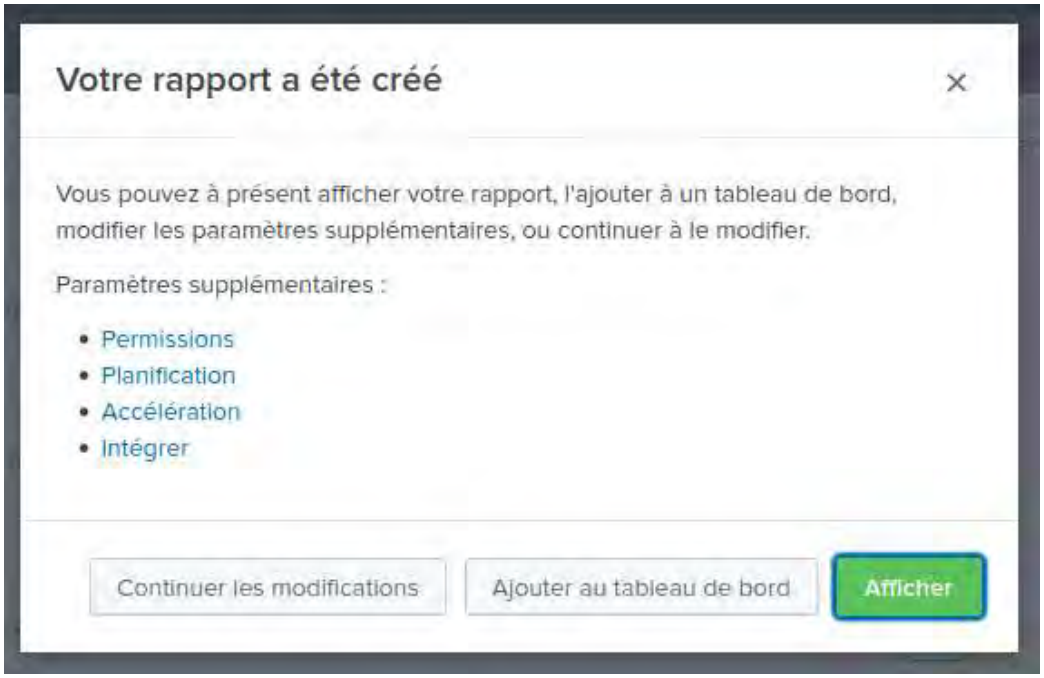


Figure 7.f : Création de rapport réussi.

Si nous cliquons sur Afficher à l'étape ci-dessus, nous pouvons voir le rapport. Nous obtenons également des options de configuration après la création du rapport.

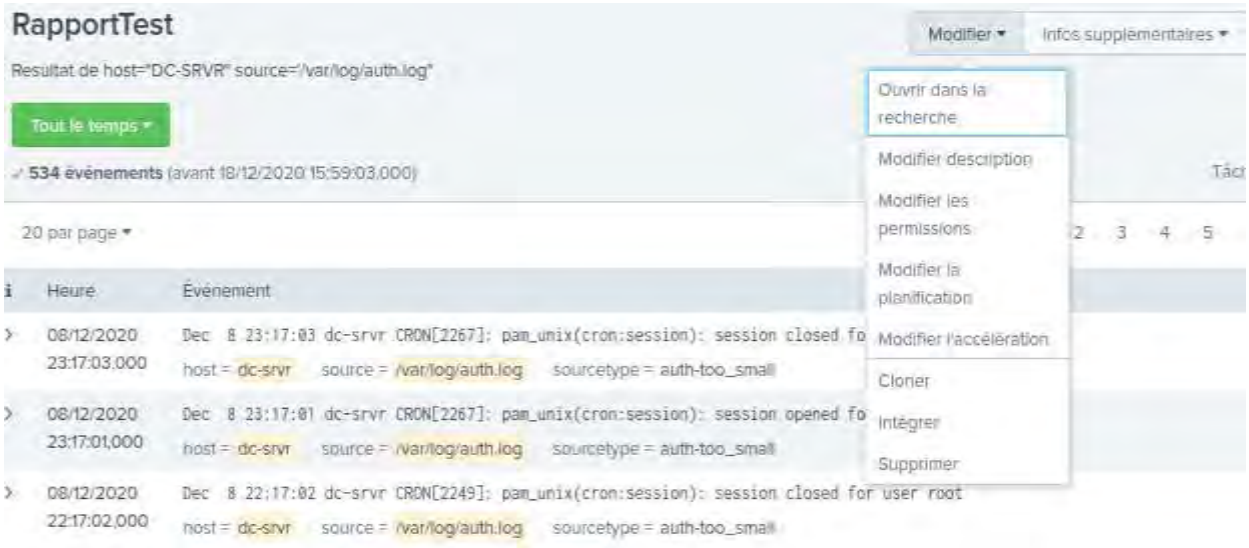


Figure 7.g: Affichage du rapport crée.

L'option **Modifier** nous permet de modifier les autorisations, le calendrier, etc., nous pourrions aussi modifier la chaîne de recherche d'origine. Cela est possible en choisissant l'option **Ouvrir** dans la recherche comme indiqué dans l'image ci-dessus. Cela ouvrira à nouveau l'option de recherche d'origine que nous pouvons modifier pour une nouvelle recherche comme indiquer à l'image ci-dessous.

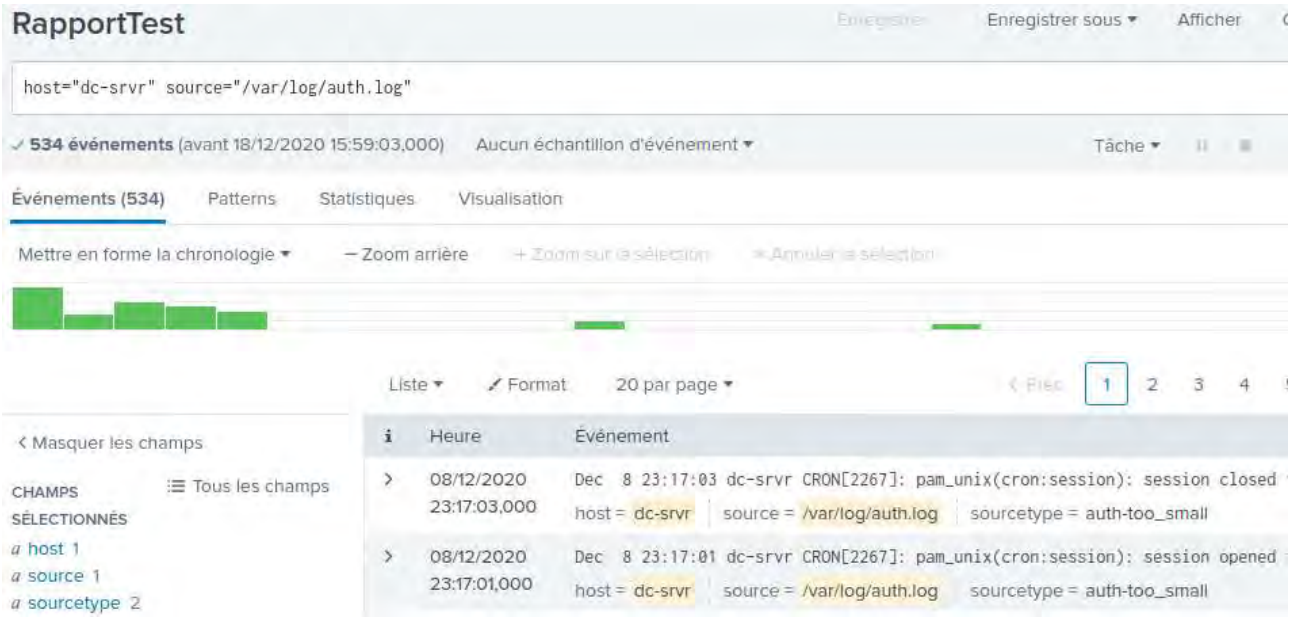


Figure 7.h : Modification de rapport.

VIII. Tests sur quelques domaines d'applications de Splunk ES

VIII-1-Exemple de surveillance avec Splunk :

Nous demandons à Splunk d'indexer (pointer) ver un répertoire /var/log/syslog sur sa propre base de données afin qu'il surveille la machine Ubuntu sur laquelle il est installé.

Créeons un index nommé Test : Parametre>index>nouveau index

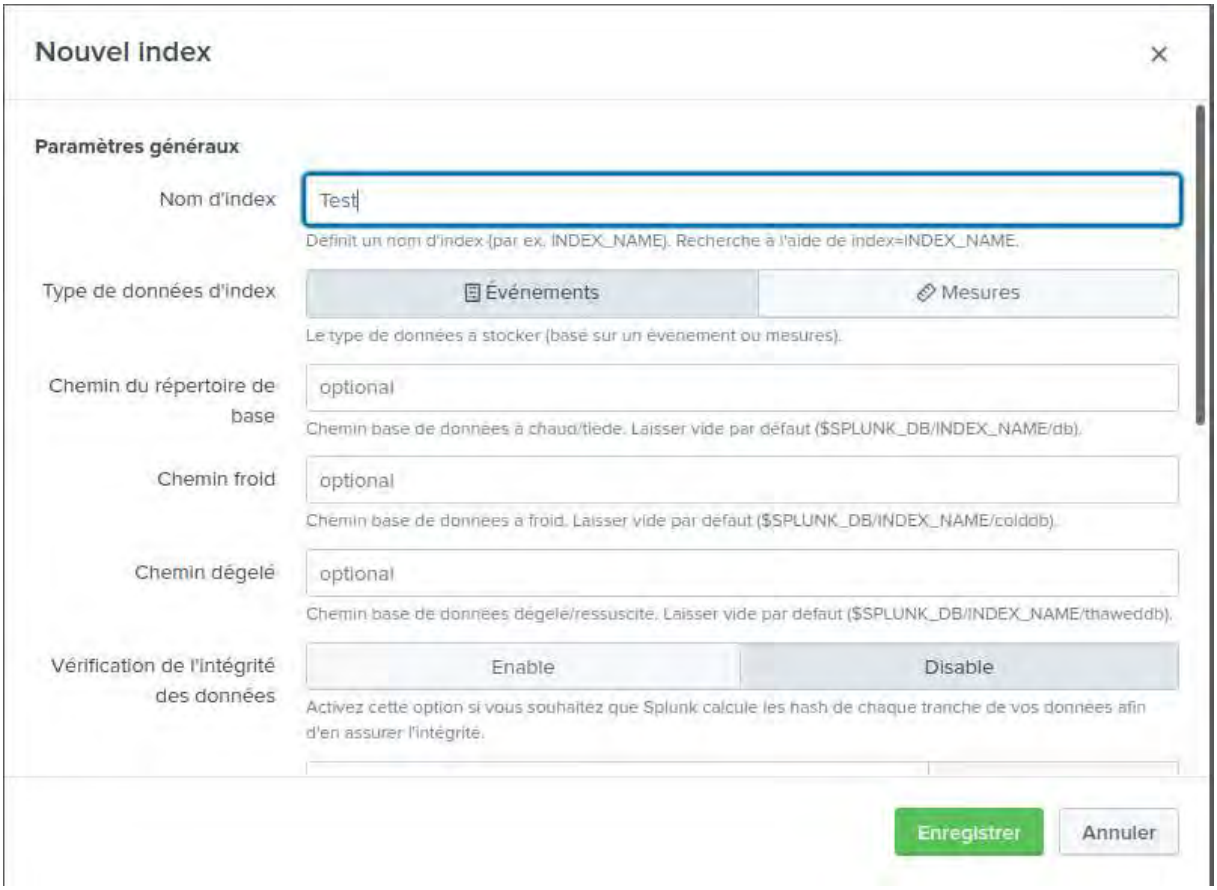


Figure 1.a : Création de l'Indexeur.

Ensuite ajouter la source de données :

Figure 1.b : Résumé.

Figure 1.c : Ajout des données.

VIII-2-Exemple de recherche de logs sur le serveur Web :

Exemple de recherche de log du serveur web sur le serveur Splunk :

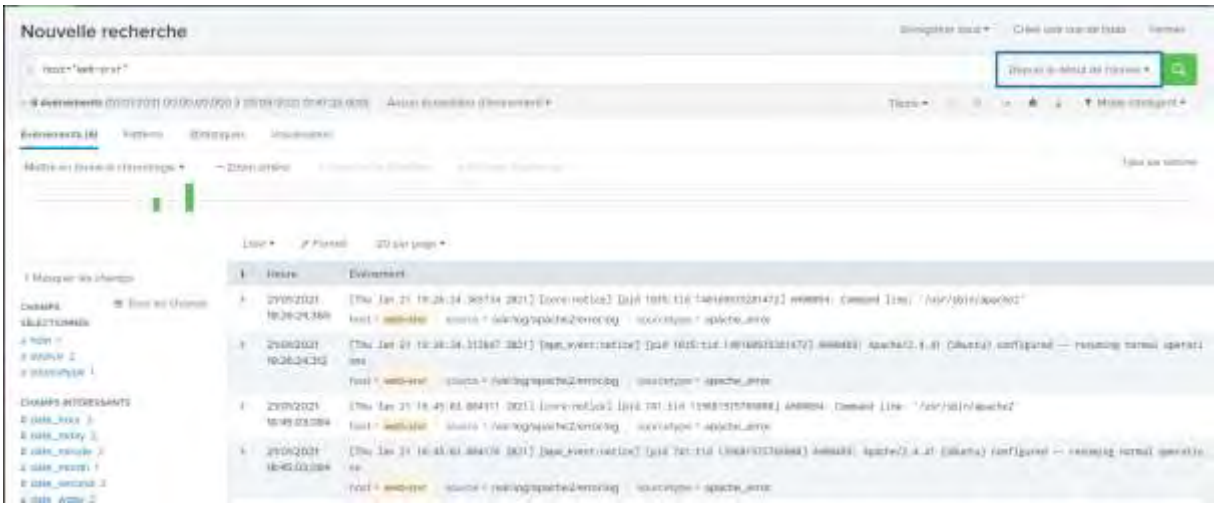


Figure 2. : Affichage des logs du serveur web.

Avec les logs du serveur web nous avons généré un tableau de bord qui nous donne une vue sur le nombre d'événements à date :

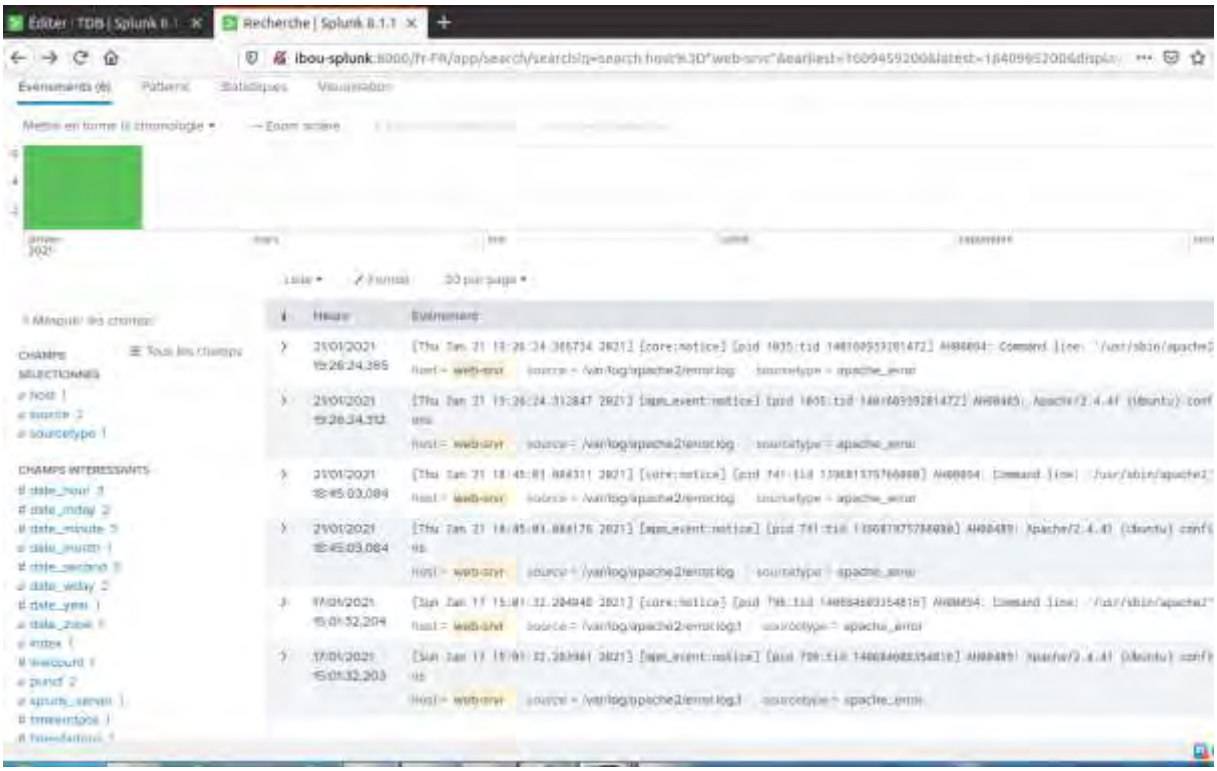


Figure 2.b : Tableau de bord.

CONCLUSION GEENERALE :

L'objectif de ce rapport était d'apporter des réponses quant au fonctionnement des SIEM dans le cadre d'une gestion centralisée. Pour cela, nous avons commencé par expliquer le contexte : l'intérêt de l'utilisation de cette méthode centralisée. Nous avons ensuite étudié le fonctionnement théorique de cette méthode, ainsi qu'une implémentation concrète.

Notre étude nous a aussi permis d'étudier les principaux avantages de la gestion centralisée :

- la simplicité de configuration du manager
- la visibilité globale du réseau par le manager
- la cohérence des alertes et des décisions menées.

Cependant cette technique possède aussi des faiblesses. Par exemple la centralisation peut amener rapidement un problème de disponibilité du service si le manager est défaillant. Nécessité d'équiper les professionnels de la sécurité de l'information d'un niveau de facilité pour identifier de manière proactive des menaces à la sécurité. La conformité réglementaire est aussi considérée comme le facteur le plus important pour l'adoption de la SIEM par les organisations.

Nous pouvons conclure que les SIEM en gestion centralisée permettent une réelle sécurisation du système, mais que l'on doit s'assurer de leur disponibilité, et qu'il semble malvenu d'utiliser les envois non sécurisés de logs. Le SIEM étant l'outil central des équipes d'un SOC puisque c'est la plateforme vers laquelle convergent toutes les données générées par l'infrastructure informatique. Il réalise les corrélations d'événements afin d'identifier d'éventuelles attaques.

Les SIEMs de nouvelle génération, qui incluent de nouvelles fonctionnalités comme l'analyse comportementale, l'apprentissage automatique et l'automatisation SOC, ouvrent de nouvelles possibilités aux analystes de sécurité.

Nous comptons dans l'avenir combiner notre solution SIEM avec un SOC afin de garantir un niveau de sécurité optimal du système d'information.

BIBLIOGRAPHIE:

B1. Carson Zimmerman, Ten Strategies of a World-Class Cybersecurity Operations Center, The MITRE Corporation, 2014.

B2. Chikonga Maimbo. Exploring the Applicability of SIEM Technologyin IT Security. MCIS 2014.

B3. HAFID Oussama. Corrélation de logs et Audit des actions des administrateurs à haut privilèges. Finatech

B4. Jacques Saraydaryan. Détection d'anomalies comportementale appliquées à la vision globale. 2008-ISAL-XXX.

B5. Jean-Olivier Gerphagnon. Attaques Informatique. CBPF-NT-007/00

B6. Laurent NOE. Conception d'un projet SIEM. OikiaLog

B7. LIVRE BLANC. Architecture validée par Splunk. Splunk

B8. Zeyed Morteza Zeinal. Analysis of Security Information ant EventManagement(SIEM) evasion and detection methods. ITC70LT 2016

WEBOGRAPHIE:

W1. varonis[en ligne]. Publié le 15/06/2020 https : //www.varonis.com/blog/what-is-siem/

W2. comparitech[en ligne]. Publié le 03/08/2020 https : //www.comparitech.com/net admin/siem-tools/

W3. searchsecurity[en ligne]. Publié le 03/02/2020 :

https : //searchsecurity.techtarget.com/definition/security-information-

and-event-management-SIEM 07/04/2020

W4. csooline[en ligne]. Publié le 28/11/2017

https : //www.csoonline.com/article/2124604/what-is-siem-software-how-it-works-and-howto choose-the-right-tool.html

W5. blog[en ligne]. Publié le 24/03/2019 https : //blog.logsign.com/security-information-and event-management-architecture/ April 24, 2019

W6. exabeam[en ligne]. Publié le 05/10/2020 https : //www.exabeam.com/siem-guide/siem architecture/

W7. softwaretestinghelp[en ligne]. Publié le 02/08/2020 https : //www.softwaretestinghelp.com/siem-tools/

W8. esecurityplanet[en ligne]. Publié le 01/07/2020 https://www.esecurityplanet.com/products/top-siem-products.html

W9. medium[en ligne]. Publié le 19/05/2020 https : //medium.com/@eakbas/the-math-of-siem-comparison-91fad2_7cac

W10. docs[en ligne]. https : //docs.splunk.com/Documentation/Splunk 07/04/2020

W11. tutoralspoint[en ligne]. https : //www.tutorialspoint.com/splunk/ 01/02/2021

W12. geek-university[en ligne]. https : //geek-university.com/splunk/ 03/03/2021

W13. Link By Net https : //www.linkbynet.com/fr/recourir-a-un-siem 03/2020