Evaluation of Splunk Enterprise for Monitoring and Investigating Cyber-Attacks

Mahir Asif Khan Nargis
210777196
Dr Edmund Robinson
MSc. Computer Science

***Abstract*—Cyber security is essential because it protects sensitive private data from theft and damage. Organisations and public members are becoming accustomed to the increasing number of cyber-attacks. Cyber security incidents have slowly evolved into targeted and sophisticated threats that target organisations and affect entire nations. It is becoming increasingly challenging to secure high-value and critical assets. This paper explores the fundamental concepts required to monitor and respond to a cyber threat incident in a corporate environment. We consider the correct approach organisations need to take to investigate attacks, effectively monitoring and detecting events using SIEM solutions to reach the state of cyber resilience. Using a well-known SIEM Splunk Enterprise, we aim to evaluate a modern SIEM's operations, general capabilities and future improvement areas.**

Keywords—SIEM, Splunk, Cyber threats, Investigation.

I. INTRODUCTION

Governments around the world are bringing more attention to cybercrime. The UK adopted the General Data Protection Regulation (GDPR) in 2016, replacing the 1995 Data Protection Directive written when the internet was in its infancy. It forces organisations that operate in the EU to report any data breaches to the public, appoint a data protection officer whose role is to inform and advise the organisations on their data protection obligations, a requirement of user consent to process information and anonymity of the data for privacy. [5] This ensures that organisations are engaged and invested in protecting their user's data. The National Cyber Security Center (NCSC) was established to help UK organisations minimise harm from cyber-attacks and provide practical guidance that is updated with recent discoveries in the field. [6]

Cyber security refers to the ability of an individual or organisation to reduce the risk of a cyber-attack. This is accomplished by protecting every device, such as computers, laptops, phones, printers and any other electronic device connected to the network. The services we can access are also vulnerable to theft and damage. Internet is a fundamental part of modern life, and it isn't easy to imagine how organisations would function correctly without it. The rise in cloud computing services such as Amazon Web Services and IoT (Internet of Things) has introduced many vulnerabilities that did not exist a few decades back. [4]

Cyber defence is a computer defensive mechanism that focuses on preventing, detecting and providing timely responses to attacks and threats so that no infrastructure or information is tampered with. It also looks toward protecting sensitive data and assets. Different organisations invest different amounts into having a solid cyber defence. Quantifying the risk, impact and harm caused by a cyber-attack becomes a matter of risk assessment. [3]

In terms of cyber security, the impact tends to look on the negative side; it refers to the result of an attack and is oriented toward harm, undesirable outcomes and impairment of organisational goals. Risk assessments are also critical, no matter the organisation's size. They allow organisations to know more about themselves, the services they require to function and the data they need to protect. [3]

Cyber-harm is a broad topic of research that consists of different types of harm. Not only physical/digital harm (damaged goods, theft) but also other forms, such as economic harm (negative financial consequences), psychological harm (affected employee's mental well-being) and reputational harm (general opinion of customers/users). [3] Cybercrime global costs in 2021 were approximately $6 trillion, which could rise to $10.5 trillion annually by 2025. [7]

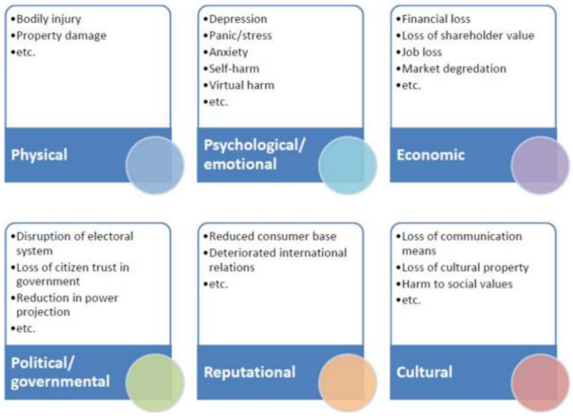


Fig.1. Classification of cyber harm and some examples.

II. COMMON CYBER THREATS

A. *Phishing*

Phishing or Social Engineering is the art of manipulating people to obtain confidential information. In cyber security, a system is as strong as its weakest link; within a secure system, the weakest link tends to be employees. [1] Cyber attackers will choose the easiest path to access a network or system. Human error is a significant access point for cyber attackers. Social engineering in the form of ransomware, trojans and spyware has become the preferred method to establish a point of entry for a cyber-attack. It is the first step of more significant system infiltration to steal sensitive data or disperse malware. This has raised the need for the introduction of security education programs which aims to bring a more robust set of protocols and transparent policies for employees when faced with a data breach or blackmail. [2]

B. *Malware*

Malware is an extensive term that refers to any software which, if executed on a system, will cause harm. Examples of malware:

- Trojans: by disguising itself as legitimate software (trojans).
- Ransomware: causing a device to become blocked or unusable, stealing, deleting or encrypting data (ransomware).
- Spyware: obtaining credentials that can access the organisation's systems and services, mining cryptocurrency in the background or passively listening and collecting data.
- Adware: displays advertising content on the user's screen and is hard to get rid of. [8]

C. *Ransomware*

Ransomware builds upon malware which has been used to block or encrypt critical sensitive data. Generally, the attackers will contact the victim organisations and demand a payment paid anonymously through cryptocurrency like Bitcoin, which is hard to trace. There is also no guarantee that access to the system or data will be returned; if the ransom is paid, it will increase the chances of being targeted. Ransomware is the most profitable malware because it relies on an unsuspecting person opening a link and is not very hard to spread. [8]

D. *Denial of Service Attack (DoS)*

Denial of Service Attack (DoS) aims to render a service inaccessible. DDoS is also a DoS attack but involves more than one vector or computer. Frequent DDoS attacks that are reported involve high-profile websites whose servers and networks are targeted with incoming requests for data. When the servers are overwhelmed trying to fulfil the recommendations, it results in a complete shutdown of the website or slows down traffic significantly as there is no bandwidth available [10].

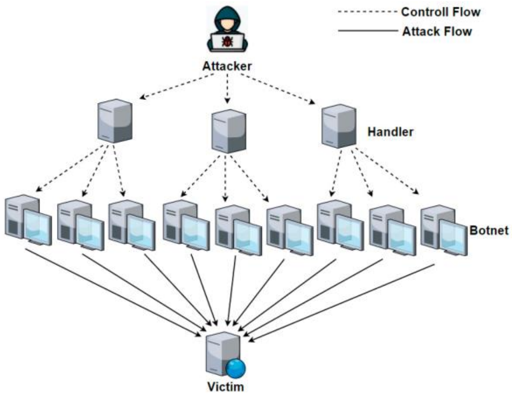


Fig.2. DDoS Attack Diagram - Notice how the attacker can use multiple devices to overload a server

E. *Zero-day Exploit*

"Zero day" is a term to describe recently discovered security vulnerabilities. The software developer requires time to work on a solution to patch the vulnerability, allowing attackers to take advantage of the vulnerability and exploit it.

Zero-day attacks cannot be prevented unless the developer patches the vulnerability before anyone notices it. Organisations need to have well-managed security solutions to limit the damage of any attack in the form of automated responses and an incident response plan everyone is aware of.

III. SIEM (SECURITY INCIDENT AND EVENT MANAGEMENT)

The National Institute of Standards and Technology defines SIEM as an application that provides the ability to gather security data from information system components and presents the data as actionable information via a single interface. In addition, SIEMs can parse logs and make them available for searching and querying.

SIEM is a critical cyber security tool for building a SOC (Security Operations Center). SOCs are responsible for detecting and responding to cyber-attacks. A SOC is established to limit the damage from cyber threats that bypass present preventive security measures. Their functions involve vulnerability management using software like Nessus or an assessment of compliance of activities and system configurations through SIEM software. [11]

SIEMs can be split into SIM (Security Information Management) and SEM (Security Event Management). A SIEM tool can be used to detect and investigate by:

- Collecting security event logs and telemetry in real-time for threat detection.
- Storing relevant events and logs for future comparison.
- Producing reports and alerts on events and suspicious activities.
- Analyze logs and network telemetry in real-time to detect events of interest.
- Investigate incidents to determine the potential severity and impact on a business

The primary data source for SIEM software has usually been time-series-based log data from sensors such as (intrusion detection systems, firewalls, antiviruses, etc.). Still, newer solutions take into account real-time data and other types of data like Active Directory, Configuration management database, vulnerability management data, HR information and threat intelligence to provide more context and information for the AI to correlate and deliver alerts on suspicious activity.

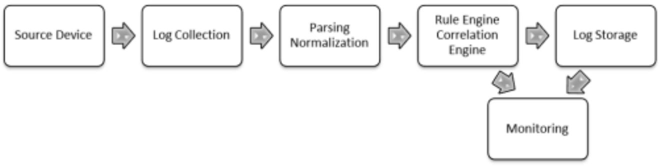


Fig. 3. Basic Structure of a SIEM

IV. LITERATURE REVIEW AND RELATED WORK

A. Background Research

The term SIEM was conceived by Mark Nicollett and Amrit Williams of Gartner in 2005 [12]. With the increase in the number of security systems, a need for centralised storage of logs was introduced to help security teams analyse data quickly and remove the requirement of access to the different types of software when performing investigations. The limiting factor for these early SIEM platforms was the lack of scalability; importing data was a hassle, and dashboards/alerts were simple. They also required increasingly expensive storage hardware to handle the increasing number of logs and reports [13].

The analysis and evaluation of SIEM systems have been widely proposed for research. Some studies focus on the commercial aspects, and others concentrate on innovative technical features that can improve current SIEM solutions. The SIEM market is very volatile and fast-changing, leading vendors like ArcSight, LogLogic and Symantec, who were leading the market in 2010, have now become unpopular due to their lack of visualisation tools and correlation rules available. Instead, newer solutions such as Splunk and Exabeam have become the consumer and organisations' primary choices. [15].

B. Visualisation Tools

In 2007, visualisation tools were extensively developed to aid in analysing data collected by security systems. While SIEM solutions brought all the logs and data together in one place, it also became impossible for a security analyst to sort through millions of log records. The research proposed creating visualisation techniques to help administrators recognise attacks through computer ports. Visualisation tools were designed for IDS systems that simplify network traffic analysis through line graphs and tables. Textual methods slowly became obsolete, and ways to scale big data were developed to fit everything into a single interface without overlap. Simple IDS alarm tools were introduced [16].

C. Big Data Analytics

As large enterprises routinely collect terabytes of data a day. They generate anywhere from ten to a hundred billion events depending on their size. To address the problems with the analysis and processing of large volumes of information, the Cloud Security Alliance (CSA) created the Big Data Working Group in 2012. The first breakthrough came from Apache Hadoop, an open-source framework that can process and store large amounts of data in multiple computers/servers instead of having one sizeable centralised server [17]. A practical module in Apache Hadoop is MapReduce, and it helps programs perform parallel computation data. Large organisations could scale horizontally by relying on lower spec hardware and cluster servers.

D. Machine Learning with SIEM

Initially, malware detection took a signature-based approach, which is very efficient and easy to implement to detect known malware already in circulation. A signature represents a series of bytes in a file or a cryptographic hash of a file. The signature from the file is then compared to a database of signatures for malicious software. The downside to this method is that it fails to recognise new variants and attack vectors which are not present in the database. [18]

Anomaly-based detection uses API call sequences, which are knowledge patterns of the expected behaviour of users or systems to look for any suspicious events. Such detection methods have many benefits as they can be implemented as real-time detectors and flag new types of malware or attacks. They do not need to rely on set rules and are more flexible in their protection. [18]

Machine learning is the foundation of anomaly-based detection methodologies. In 2015 Richard Yang and Victor Kang suggested using data mining techniques such as N-gram and PE features on sample files and using them as weights for a supervised algorithm that can classify the file as malware or benign. Then, an unsupervised algorithm will collect all the malware files and determine their severity. The methodology was evaluated with the overall accuracy score, noting the number of true positives and false positives. Their application of the Random Forest algorithm reached an overall accuracy of 94%, with a meagre number of false positives. Some improvements suggested using K-means clustering to better group different types of malware. [19]

E. Importance of SIEM

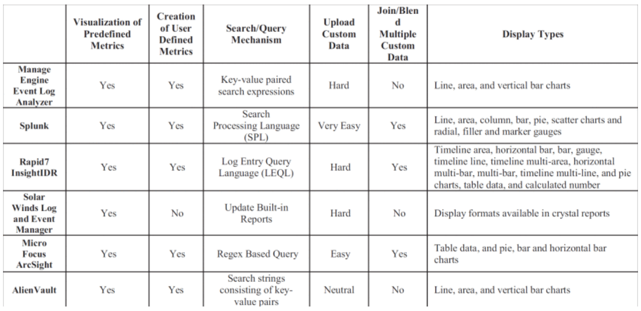
A SIEM solution aims to identify, protect, detect and respond to security-related incidents. The conference paper by Oskars and Andrejs looks at the advantages of a SIEM for an enterprise or business that requires a resilient, secure network and complete security audits. Although an expensive solution, enterprises still decide to invest in SIEM solutions due to the positive net outcome from successfully thwarting cyber-attacks. [20]

F. SIEM for Malware detection

The 2021 conference paper by Marian implements various security rules on Splunk to monitor a system in real-time and create alerts when the presence of Mirai IoT malware is detected. This briefly introduces how Splunk functions, a summary of Splunk's forwarders and the capabilities of Splunk's IP reputation lookup and its query language SPL. [21]

G. Evaluation of SIEM solutions

Gartner, Inc is a technological research and consulting firm that evaluates SIEM solutions. The SIEM products are divided into four categories: Leaders, Challengers, Visionaries and Niche players. The parameters considered for a SIEM are its execution (services, viability, sales and pricing) and vision (marketing strategy, sales strategy, business model, innovation and geographic strategy). [22]



	Visualization of Predefined Metrics	Creation of User Defined Metrics	Search/Query Mechanism	Upload Custom Data	Join/Blend Multiple Custom Data	Display Types
Manage Engine Event Log Analyzer	Yes	Yes	Key-value paired search expressions	Hard	No	Line, area, and vertical bar charts
Splunk	Yes	Yes	Search Processing Language (SPL)	Very Easy	Yes	Line, area, column, bar, pie, scatter charts and radial, filler and marker gauges
Rapid7 InsightIDR	Yes	Yes	Log Entry Query Language (LEQL)	Hard	Yes	Timeline area, horizontal bar, bar, gauge, timeline line, timeline multi-area, horizontal multi-bar, multi-bar, timeline multi-line, and pie charts, table data, and calculated number
Solar Winds Log and Event Manager	Yes	No	Update Built-in Reports	Hard	No	Display formats available in crystal reports
Micro Focus ArcSight	Yes	Yes	Regex Based Query	Easy	Yes	Table data, and pie, bar and horizontal bar charts
AlienVault	Yes	Yes	Search strings consisting of key-value pairs	Neutral	No	Line, area, and vertical bar charts

Fig.4. Evaluation summary of popular SIEM vendors and their characteristics.

H. The Diamond Model of Intrusion Analysis

The diamond model of intrusion analysis breaks down an incident into four components: adversary, infrastructure, capability and victim. Released in 2006, it established a formal method of applying scientific principles to the field of intrusion analysis. [31] Before this time, security analysts often relied on data print-outs and intuition to conduct threat hunting and investigations. It analyses the motives and documents the thought process of the adversary. It provides a framework for investigating any intrusion and is still followed today.

V. METHODS AND DATASET

A. Splunk Inc.

Splunk is an American software company based in California which produces SIEM solutions for searching, monitoring and analysing machine-generated data through a simple web interface. Machine-generated data is complex to understand for the human eye and is, in most cases, unstructured. Not much information can be gained, and time is wasted on analysing such data logs. As of September 2020, Splunk saw a massive rise in revenue, and Splunk's client list included 92 companies on the Fortune 100 list [23].

Splunk's premium software consists of three main components: the search head, the indexer and the forwarders.

- Indexer: Raw machine-generated data and logs are processed and stored in directories with indexes to aid information retrieval and to query.
- Search Head: It allows for querying for different events. Splunk uses its own SPL (Search Processing Language) query language. It can be used for data searching, filtering, modification, insertion and deletion. Syntax was based on Unix Pipeline and SQL, but it's optimised for time series data.

- Forwarders: Installed in the network's machines, collects data/logs, IDS and other firewall information, which are sent to the indexer. Splunk Universal Forwarder runs in the background and uses minimal resources compared to other SIEM solutions.

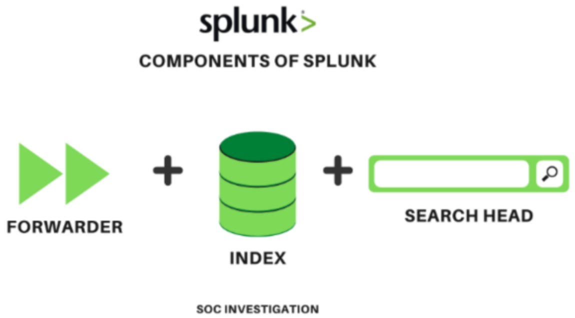


Fig.5. Splunk's Three Main Components [14]

Splunk has introduced the use of add-ons that improve its SIEM capabilities. Two special add-ons are UEBA and SOAR. These two technologies are critical in aiding the detection of anomalies, automation of actions and setting up alerts. This is what sets Splunk from other outdated SIEM solutions.

B. UEBA

UEBA stands for User and Entity Behaviour Analysis and is a combination of algorithms and machine learning used to detect anomalies of users and any equipment such as routers, servers and endpoints. It detects any irregularities from standard everyday patterns, such as a user who suddenly connects using a VPN from Ukraine when his normal usage of VPN is connecting from the UK. This would flag and alert the IT administrators and automatically disconnect the user. UEBA is very good at detecting compromised accounts and unexpected traffic. Its goal is to reduce the IT analysts required and the organisation's costs [24].

C. SOAR

SOAR (Security Orchestration Automation Response) is a combination of software created to automate the duties of security analysts and reduce the time needed to respond to security incidents. In the event of malware detection, the IT administrator would need to connect to the endpoint security to quarantine the computer and use an IDS to search for the source of the malware. SOAR allows for the automation of this process through Splunk's dashboard [25].

D. SPL

SPL (Search Processing Language) has over 140+ search commands, including anomaly detection and machine learning functions. SPL is a powerful tool that can be compared to a web search engine but for logs. There are simple searches such as a username or IP address which will present specific records for that machine or user. More complex searches are also possible, such as finding out what applications are causing a delay on startup and then presenting the results as a graph [26].

The SPL basic structure is as follows:
Search and filter | munge | report | cleanup

The full SPL query for Identifying transactions and anomalies:
Sourcetype=dataset_name* | transactions SESSIONID |stats min(Duration) max (Duration) avg (Duration) |Inputlookup data.csv |anomalydetection action=summary | method=iqr action=remove

Group by session ID (Search and filter of transactions by their SESSIONID):
Sourcetype=dataset_name* | transactions SESSIONID

Calculate Session Duration:
|stats min(Duration) max (Duration) avg (Duration)

Input a dataset with common information about transactions and use the anomaly detection function to detect the anomalies. Finish off by summarising this information and using IQR to remove outliers:
|Inputlookup data.csv |anomalydetection action=summary | method=iqr action=remove

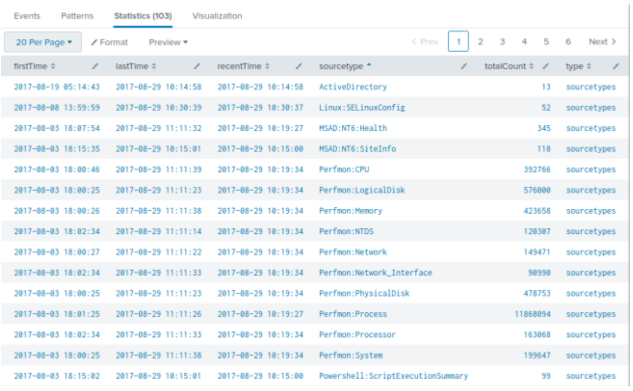
E. Boss of the SOC Version 2

BOTSv2 Dataset was created by the Splunk Security specialist team, which contains data from a few laboratory environments posing as a Beer company named Frothly and connected to the internet. The Splunk Heavy forwarders used best practices for Windows endpoint monitoring, Microsoft Sysmon deployment and Windows Event logging [27].

To perform the initial information gathering on the dataset, we will use the following metadata command:

- **metadata type=sourcetypes index=botsv2 | eval firstTime=strftime(firstTime,"%Y-%m-%d %H:%M:%S") | eval lastTime=strftime(lastTime,"%Y-%m-%d %H:%M:%S") | eval recentTime=strftime(recentTime,"%Y-%m-%d %H:%M:%S") | sort – totalCount**

Splunk assigns specific fields to the indexed data. The more common fields used are time (first and last event), source (file path, protocol or port value), source type (type of metadata) and host (hostname or IP).



firstTime	lastTime	recentTime	sourcetype	totalCount	type
2017-08-19 05:14:43	2017-08-29 10:14:58	2017-08-29 10:14:58	ActiveDirectory	13	sourcetypes
2017-08-08 13:59:59	2017-08-29 10:30:39	2017-08-29 10:30:37	Linux:SELinuxConfig	52	sourcetypes
2017-08-03 18:07:54	2017-08-29 11:11:32	2017-08-29 10:19:27	MSAD:NT6:Health	345	sourcetypes
2017-08-03 18:15:35	2017-08-29 10:15:01	2017-08-29 10:15:00	MSAD:NT6:SiteInfo	118	sourcetypes
2017-08-03 18:00:46	2017-08-29 11:11:39	2017-08-29 10:19:34	Perfmon:CPU	392766	sourcetypes
2017-08-03 18:00:25	2017-08-29 11:11:23	2017-08-29 10:19:34	Perfmon:LogicalDisk	576000	sourcetypes
2017-08-03 18:00:26	2017-08-29 11:11:38	2017-08-29 10:19:34	Perfmon:Memory	423658	sourcetypes
2017-08-03 18:02:34	2017-08-29 11:11:14	2017-08-29 10:19:34	Perfmon:NTDS	120307	sourcetypes
2017-08-03 18:00:27	2017-08-29 11:11:22	2017-08-29 10:19:34	Perfmon:Network	149471	sourcetypes
2017-08-03 18:02:34	2017-08-29 11:11:33	2017-08-29 10:19:34	Perfmon:Network_Interface	90990	sourcetypes
2017-08-03 18:00:25	2017-08-29 11:11:23	2017-08-29 10:19:34	Perfmon:PhysicalDisk	478753	sourcetypes
2017-08-03 18:01:25	2017-08-29 11:11:26	2017-08-29 10:19:27	Perfmon:Process	11868094	sourcetypes
2017-08-03 18:02:34	2017-08-29 11:11:33	2017-08-29 10:19:34	Perfmon:Processor	163068	sourcetypes
2017-08-03 18:00:25	2017-08-29 11:11:38	2017-08-29 10:19:34	Perfmon:System	199647	sourcetypes
2017-08-03 18:15:02	2017-08-29 10:15:01	2017-08-29 10:15:00	Powershell:ScriptExecutionSummary	99	sourcetypes

Fig.6. Metadata search on the BOTSv2 dataset. The metadata command provides information on when the first/last event occurred and the count of the total number of events for each source type [28].

Boss of the SOC Version 2 contains four scenarios and a list of questions made for academics, students and employees trying to learn more about Splunk. For this dissertation, we will complete the web activity investigation scenario and the USB attack investigation. A further step-by-step guide on the path I followed to solve the questions is provided in the appendix.

VI. INVESTIGATIONS, ANALYSIS AND RESULTS

Using Splunk, we will be conducting investigations on two incidents involving web activity and ransomware. The aim is to implement a successful incident response, where we carry out reconnaissance (harvesting for information, email addresses, IPs involved, etc.) and determine the payload or malware and mode of delivery. Other things to note are determining if the threat is still active, removing any malware from the systems and identifying the intruder's original goals.

A. Web Activity Investigation

The investigation will examine the web activity of a Frothly employee named Amber Turing. Amber Turing was accused of contacting a competitor's CEO and providing them with internal information about Frothly. We aim to find evidence and shed light on the matter by going over Amber Turing's machine logs.

The following SPL commands were used:

- **index="botsv2" sourcetype="pan:traffic" amber**
- **index="botsv2" 10.0.2.101 sourcetype="stream:http" | dedup site | table site**
- **index="botsv2" sourcetype="stream:http" berkbeer.com**
- **index="botsv2" sourcetype="stream:smtp" amber**
- **index="botsv2" sourcetype="stream:smtp" aturing@froth.ly berkbeer.com**

To find what websites Amber has accessed, we need to find the IP Address of her system. IP addresses can be found in Splunk's PAN source type. With Amber's IP, we can examine her HTTP traffic logs which should provide us with websites she has visited. There are 107 websites produced after removing duplicates, and thus we are required to further filter for the industry and ignore Bing/Microsoft website traffic.

After finding the competitor's website, we can narrow Amber's HTTP traffic to the website and see what resources she accessed. We want to know if she obtained the CEO's name and if we can find the email address Amber contacted. Once we obtain the CEO's email address, we can build a search query involving SMTP (Simple Mail Transfer Protocol) to view the email traffic between Amber and the CEO.

By inspecting the emails sent by Amber to Mr. Martin Berk (CEO of Berkbeer), we can see that Amber was not happy about her job at Frothly and wanted to build her connections with Berkbeer. From the conversations, an employee from Berkbeer asked Amber for confidential information. Going through the raw data of the last event, we can see she sent an attachment with information on a patent from Frothly.

B. USB Ransomware Attack Investigation

The scenario involves an individual named Mallory, her Mac OS machine and some encrypted files. The attack occurred due to an infected USB ransomware attack, and she reported that some important PowerPoint files were encrypted and unrecoverable. This investigation aims to find information about the USB used in the attack and information about the malware.

The following SPL commands were used:

- **index="botsv2" Mallory | table host**
- **index="botsv2" host="MACLORY-AIR13" (*.ppt OR *.pptx)**
- **index="botsv2" host="MACLORY-AIR13" (*.crypt) sourcetype=ps**
- **index="botsv2" host=kutekitten sourcetype=osquery_results "columns.target_path"="/Users/mkraeusen/Downloads/Important_HR_INFO_for_mkraeusen"**
- **index="botsv2" host=kutekitten USB**

To determine Mallory's Mac device name was "MACLORY-AIR13", we carried out a simple search with her name and the host index. We then investigated one PowerPoint file she reported to have been encrypted by the malware and determined the malware encrypted files contain the extension ".crypt". After searching her device for files ending with the same extension, it was found that 1022 files were corrupted.

The next step is to identify the USB device used. We know the perpetrator is an employee named Kevin Lagerfield and that the USB drive was connected to Mallory's personal MacBook "kutekitten". When Mallory ran the malware, it obfuscated itself during the execution.

There are two options to find information about the USB: looking through the registry logs in the Windows Event Logs for the event of a USB being connected to the device and using the estimated time of infection or finding the ransomware and then working out how and when the ransomware was introduced.

In order to find the malware, we will be using OSquery. OSquery exposes an operating system as a high-performance relational database. This allows you to write SQL queries to explore the operating system's data. With OSquery, SQL tables represent abstract concepts such as running processes, loaded kernel modules, open network connections, browser plugins, hardware events or file hashes [29].

Through OSquery, we can use Splunk to search for any new files downloaded into the system. Most USB attacks work by downloading software through a command. We inspected various directories and found the suspected malware file. We ran a virustotal search with the md5 obtained from the file that confirmed our hypothesis. To get information on the USB, we used some time correlation where we searched for hardware events with the keyword USB during a timeframe of 60seconds before the malware was downloaded. This allowed us to obtain the event of the USB being connected to the device and details about the USB vendor ID and model ID. We can search for further information about the USB vendor through the device hunt website [30].

VII. CONCLUSION

SIEMs faced many technical challenges in 2014; however, they were still an innate part of a SOC and have been developed to deal with event collection, storage, analysis and visualisation challenges. Splunk is one of the leading vendors for SIEM due to solving these issues and being an innovator in the market.

A. Data Input

Splunk does not work as a database, as there is no need to define tables or fields to store data. It makes data input simple and fast since it can accept every type of log. It has automatic indexing for the input data, and events are timestamped, which makes creating queries for logs extraordinarily satisfying and hassle-free. However, during my investigations, there were still a few instances where I was required to go through machine-generated data to obtain the information I was looking for.

B. Scalability

Since Splunk stores data directly in a file system, there is no database to manage and requires no prior set-up. Splunk is available for most platforms and is very easy to install on Windows machines. If one server is not enough to store all the data, it's possible to add another one with the data being distributed between the two servers evenly, which does not affect the search speed. There is also the option to store data for much longer than other SIEM software, which allows for longer data retention. Longer data retention is beneficial for training machine-learning algorithms and detecting slow-acting threats.

C. Visualisation

Splunk's machine-learning capabilities recommend fields of interest that can aid and guide new users and learners in their analysis. It displays critical areas of interest for further inspection and possible filters which can simplify the intrusion analysis.

Similarly, I took advantage of the many excellent, neat functions that can be used to visualise data and search for patterns of suspicious activity. I would have otherwise missed these patterns when sifting through raw data.

D. Alerts

Splunk allows the user to set up a trigger that returns results when it meets defined conditions or rules. This monitors real-time data or can be scheduled to run at different time intervals. The alerts can be considered previously saved searches and allow a faster response time to incidents such as zero-day attacks. It can significantly minimise the work of a security analyst to have a list of alerts set up those monitor areas of interest and flag up suspicious activity. Some companies and academics share their rules and search queries through a standard format for describing log events called Sigma. The community can suggest improvements to known rules or create new ones to detect unknown attack vectors. There is no issue with using different SIEM products as there are tools available which convert Sigma into the specific SIEM rules for Splunk, ArcSight, Azure Sentinel, etc. [32]

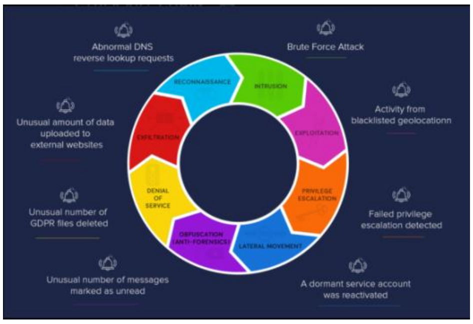

Abnormal DNS reverse lookup requests
Brute Force Attack
RECONNAISSANCE
INTRUSION
Unusual amount of data uploaded to external websites
Activity from blacklisted geolocationn
EXFILTRATION
EXPLOITATION
DENIAL OF SERVICE
PRIVILEGE ESCALATION
Unusual number of GDPR files deleted
Failed privilege escalation detected
OBFUSCATION (ANTI-FORENSICS)
LATERAL MOVEMENT
Unusual number of messages marked as unread
A dormant service account was reactivated

Fig.7. Examples of Alert Setups to Monitor Suspicious Activities

This feature of Splunk is not available in the free version, and setting it up requires a team of professionals who analyse the needs of the company and the capabilities of their adversaries or competitors. This can be very expensive for smaller companies and organisations.

VIII. LIMITATIONS AND FURTHER WORK

Gartner named Splunk a leader in the Gartner Magic Quadrant for Security Information and Event Management (SIEM) due to its ease of use and popularity among the Fortune 100. The closest competitor, IBM, was mentioned to have a better “Completeness of Vision” and has placed more effort in innovating the market.

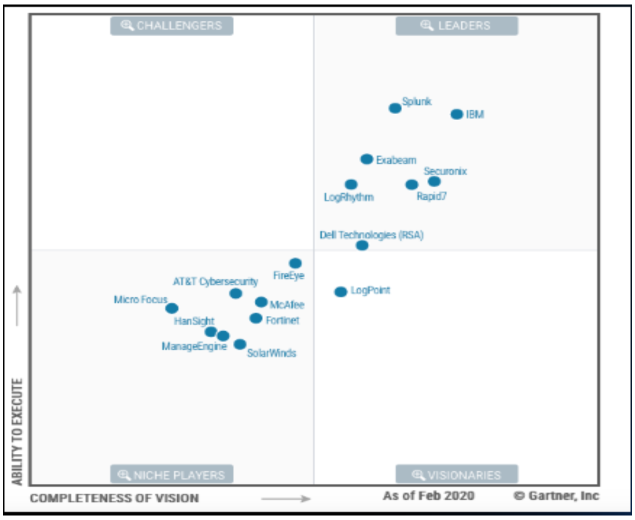


Fig.8. Gartner Inc. Magic Quadrant for SIEM solutions in Feb 2020. [33]

Although SIEM solutions have solved many of its prior problems and become more efficient over time, attackers and their attack vectors have strengthened their methods. The speed with which attackers complete the kill-chain process is increasingly faster and harder to detect. This increased risk of attack is causing insurance premiums for cyber harm to increase yearly. Further work can be done to improve the available SIEM solutions to counterattack the issue.

A. Machine Learning

Real-time detection is reliant on predefined rules and signatures. Artificial intelligence and machine learning algorithms are trained with available data. Many algorithms can detect anomalies or outliers in data, but the most popular ones currently are the SVM and Random Forest algorithms. These algorithms, however, are not perfect and have certain drawbacks. One such drawback is the inability to react to new events, and zero-day attacks, as the algorithms cannot detect patterns it has not been trained with.

Further work can be done to improve machine-learning algorithms to reduce the number of false positives. A high number of false positives (events flagged as suspicious incorrectly) can lead to alert fatigue, as it requires a security analyst to investigate the cause of the alert and look for the cause.

B. Active defence

Organisations are too passive, and waiting for sophisticated attacks to occur before reacting is not the correct solution. A system that does not rely on human input needs to be developed and should be able to anticipate attacks, respond in real-time, create traps to contain malware, and protect valuable assets from experienced hackers. It should be more challenging for the attacker to succeed than the defender.

REFERENCES

Book References:

[1] Bruce Schneier, “Secrets and Lies: Digital Security in a Networked World,” ch. 2, pp. 255-271, 2000. URL:https://www.accord.edu.so/web/content/32538?download=true&access_token=d5a8988c-4ac7-4eed-872f-c206fd8bd147 [Accessed 15 August 2022]

Report/Journal References:

[2] Hussain Aldawood and Geoffrey Skinner, “Educating and Raising Awareness on Cyber Security Social Engineering: A Literature Review,” 2018. URL:https://ieeexplore.ieee.org/abstract/document/8615162?casa_token=32TjuYXaYKcAAAAA:Ho0TRcI_NywqoHnMeuXCOoLGblLXlKHJ0jR4zI_XtP64p1Qf8d89Ic1c8-8tQcCsH6pqXzkP [Accessed 15 August 2022]

[3] Jason R C Nurse, Michael Goldsmith, Sadie Creese and David Upton, “A Taxonomy of Cyber-Harms: Defining the Impacts of Cyber-Attacks and Understanding how they Propagate,” vol. 4, iss. 1, 2018. URL:https://academic.oup.com/cybersecurity/article/4/1/tyy006/5133288?searchresult=1 [Accessed 15 August 2022]

[12] Amrit T. Williams and Mark Nicolett,”Improve IT Security with Vulnerability Management,” 2005. URL:https://www.gartner.com/en/documents/480703 [Accessed 15 August 2022]

[15] Gustavo González-Granadillo, Susana González-Zarzosa and Rodrigo Diaz,”Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures,” 2021. URL:https://pdfs.semanticscholar.org/a669/59b963c271ca3438dc516e0766317be1811c.pdf [Accessed 15 August 2022]

[16] K. Abdullah,"Scaling and Visualizing Network Data to Facilitate in Intrusion Detection Tasks," 2006. URL:https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.92.1121&rep=rep1&type=pdf [Accessed 15 August 2022]

[17] Alvaro A. Cardenas, Pratyusa K. Manadhata and Sreeranga P. Rajan,"Big Data Analytics for Security," 2013. URL:https://ieeexplore.ieee.org/document/6682971 [Accessed 15 August 2022]

[18] Mozammel Chowdhury, Azizur Rahman and Rafiqul Islam,"Protecting Data from Malware Threats using Machine Learning Technique," 2017. URL:https://ieeexplore.ieee.org/document/8283111 [Accessed 15 August 2022]

[19] Richard Yang, Victor Kang, Sami Albouq and Mohamed Zohdy,"Application of Hybrid Machine Learning to Detect and Remove Malware," 2015. URL:https://www.researchgate.net/publication/282389127_Application_of_Hybrid_Machine_Learning_to_Detect_and_Remove_Malware [Accessed 15 August 2022]

[20] Oskars Podzins and Andrejs Romanovs,"Why SIEM is Irreplaceable in a Secure IT Environment," 2019. URL:https://ieeexplore.ieee.org/document/8732173 [Accessed 15 August 2022]

[21] Marian Hristov, Maria Nenova, Georgi Iliev and Dimiter Avresky,"Integration of Splunk Enterprise SIEM for DDoS Attack Detection in IoT," 2021. URL:https://ieeexplore.ieee.org/document/9685977 [Accessed 15 August 2022]

[22] Ferda Özdemir Sönmez and Banu Günel,"Evaluation of Security Information and Event Management Systems for Custom Security Visualization Generation," 2018. URL:https://ieeexplore.ieee.org/document/8625291 [Accessed 15 August 2022]

[31]Sergio Caltagirone, Andrew Pendergast and Christopher Betz,"The Diamond Model of Intrusion Analysis," 2006. URl:http://www.activeresponse.org/wp-content/uploads/2013/07/diamond.pdf [Accessed 15 August 2022]

Web References:

[4] Anna Katrenko and Elena Semeniak, "Internet of Things (IoT) Security: Challenges and Best Practices," 2022. URL:https://www.apriorit.com/dev-blog/513-iot-security [Accessed 15 August 2022]

[5] The European Parliament and the Council of the European Union, "General Data Protection Regulation," 2016. URL:https://gdpr-info.eu/art-1-gdpr/ [Accessed 15 August 2022]

[6] National Cyber Security Centre, "About the NCSC-What We Do," 2016. URL:https://www.ncsc.gov.uk/section/about-ncsc/what-we-do [Accessed 15 August 2022]

[7] Steve Morgan, "Cybercrime to Cost the World $10.5 Trillion Annually by 2025," 2020. URL:https://cybersecurityventures.com/cybercrime-damages-6-trillion-by-2021/ [Accessed 15 August 2022]

[8] National Cyber Security Centre,"Mitigating Malware and Ransomware Attacks," 2020. URL:https://www.ncsc.gov.uk/guidance/mitigating-malware-and-ransomware-attacks [Accessed 15 August 2022]

[9]Shruti M, "10 Types of Cyber Attacks you Should be Aware in 2022," 2022. URL:https://www.simplilearn.com/tutorials/cyber-security-tutorial/types-of-cyber-attacks [Accessed 15 August 2022]

[10] National Cyber Security Centre, "Denial of Service (DoS) Guidance," 2016. URL:https://www.ncsc.gov.uk/collection/denial-service-dos-guidance-collection [Accessed 15 August 2022]

[11] National Cyber Security Centre,"Building a Security Operations Centre (SOC)," 2022. URL:https://www.ncsc.gov.uk/collection/building-a-security-operations-centre [Accessed 15 August 2022]

[13] Stephen Gailey,"A Brief History of SIEM," 2020 URL:https://cybersecurity-magazine.com/a-brief-history-of-siem/ [Accessed 15 August 2022]

[14] SOC Investigation "Splunk Architecture" 2022. URL: https://www.socinvestigation.com/splunk-architecture-forwarder-indexer-and-search-head/ [Accessed 15 August 2022]

[23] Mike Wheatley,"Data Analytics Firm Splunk sees 81% Jump in Quarterly Cloud Software Bookings," 2020. URL:https://siliconangle.com/2020/05/21/splunk-sees-81-jump-quarterly-cloud-software-bookings/ [Accessed 15 August 2022]

[24] Fortinet,"User and Entity Behavior Analytics (UEBA)," 2022 URL:https://www.fortinet.com/resources/cyberglossary/what-is-ueba [Accessed 15 August 2022]

[25] Gartner,”Security Orchestration, Automation and Response (SOAR),” URL:https://www.gartner.com/en/information-technology/glossary/security-orchestration-automation-response-soar [Accessed 15 August 2022]

[26] Splunk Enterprise,”Command Quick Reference,” 2022. URL:https://docs.splunk.com/Documentation/Splunk/9.0.0/SearchReference/ListOfSearchCommands [Accessed 15 August 2022]

[27] Splunk,”Botsv2,” 2018. URL:https://github.com/splunk/botsv2 [Accessed 15 August 2022]

[28] Splunk Enterprise,”Metadata,” 2022. URL:https://docs.splunk.com/Documentation/Splunk/latest/SearchReference/Metadata [Accessed 15 August 2022]

[29] OSQUERY,”OSQUERY,” 2016. URL:https://github.com/osquery/osquery [Accessed 15 August 2022]

[30]Device Hunt,”Device Hunt – Find your Device & Driver,” 2022. URL:https://devicehunt.com/ [Accessed 15 August 2022]

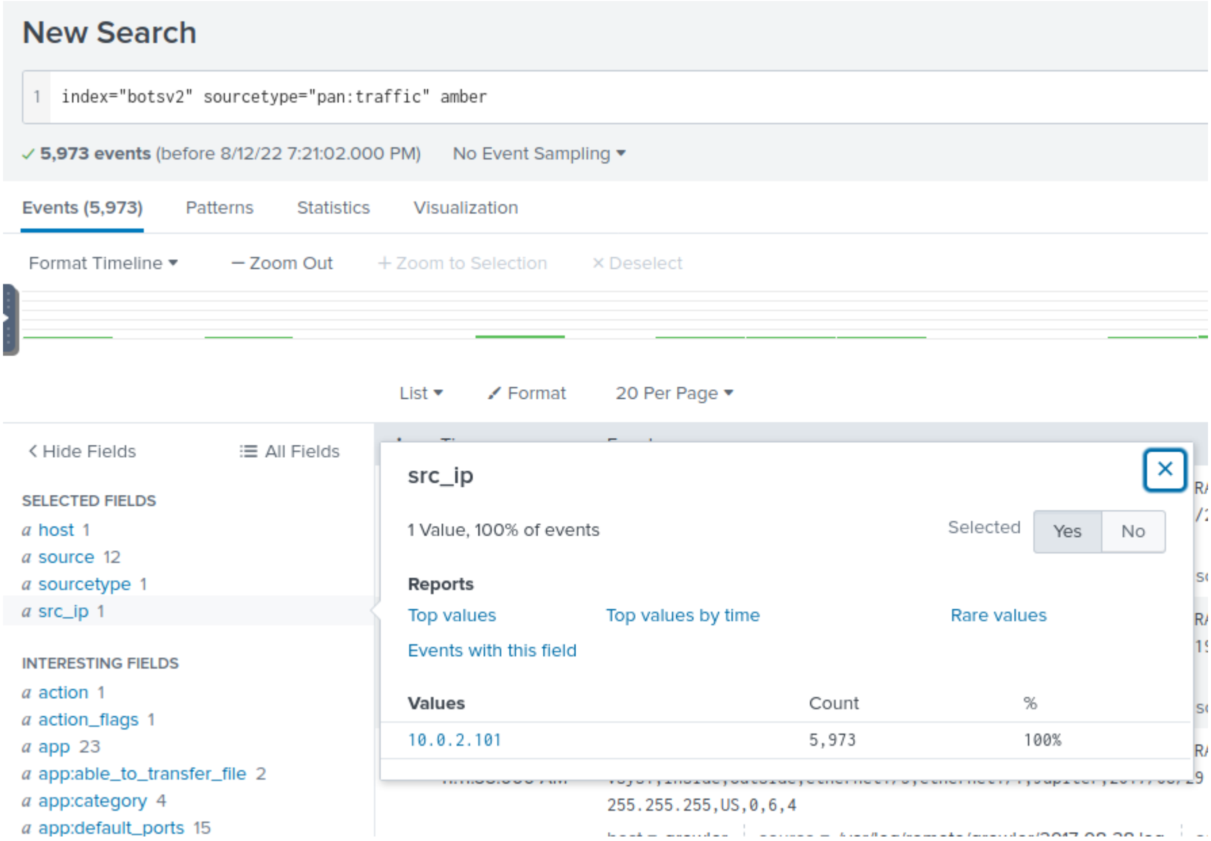
[32]SigmaHQ,”Sigma,” 2018. URL:https://github.com/SigmaHQ/sigma [Accessed 15 August 2022]

[33]Gartner,”Gartner Magic Quadrant for Security Information and Event Management,” 2021. URL:https://www.gartner.com/en/documents/4003080 [Accessed 15 August 2022]

APPENDIX

WEB ACTIVITY INVESTIGATION

index="botsv2" sourcetype="pan:traffic" amber



```
1  index="botsv2" sourcetype="pan:traffic" amber
```

✓ 5,973 events (before 8/12/22 7:21:02.000 PM) No Event Sampling ▾

SELECTED FIELDS
- a host 1
- a source 12
- a sourcetype 1
- a src_ip 1

INTERESTING FIELDS
- a action 1
- a action_flags 1
- a app 23
- a app:able_to_transfer_file 2
- a app:category 4
- a app:default_ports 15

src_ip

1 Value, 100% of events

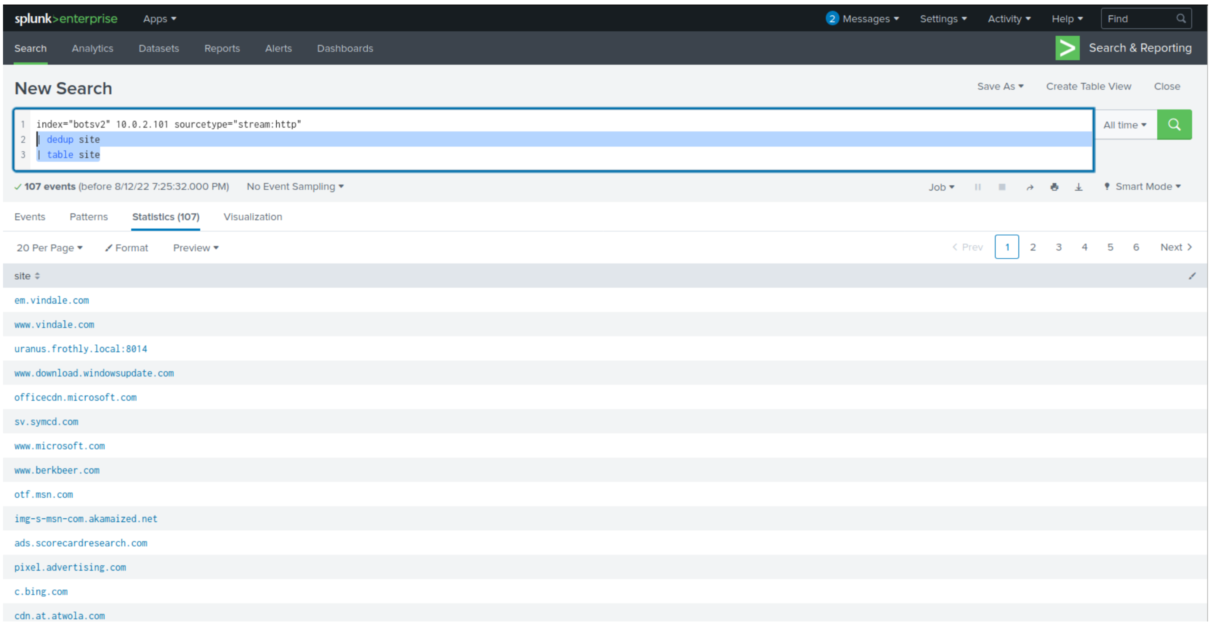
Selected [✓ Yes] No

Values	Count	%
10.0.2.101	5,973	100%

255.255.255,US,0,6,4

Fig.9. Amber’s IP is 10.0.2.101.

index="botsv2" 10.0.2.101 sourcetype="stream:http" | dedup site | table site

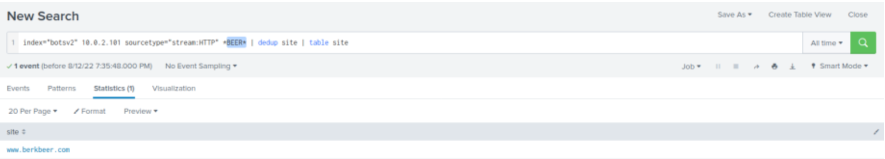


```
1  index="botsv2" 10.0.2.101 sourcetype="stream:http"
2  | dedup site
3  | table site
```

✓ 107 events (before 8/12/22 7:25:32.000 PM) No Event Sampling ▾

site
em.vindale.com
www.vindale.com
uranus.frothly.local:8014
www.download.windowsupdate.com
officecdn.microsoft.com
sv.symcd.com
www.microsoft.com
www.berkbeer.com
otf.msn.com
img-s-msn-com.akamaized.net
ads.scorecardresearch.com
pixel.advertising.com
c.bing.com
cdn.at.atwola.com

Fig.10. Search for HTTP traffic with Amber’s IP. We obtained 107 websites after removing the duplicates and displaying the results in a table.



New Search

Save As ▾ Create Table View Close

```
1 index="botsv2" 10.0.2.101 sourcetype="stream:HTTP" *BEER* | dedup site | table site
```

All time ▾

✓ 1 event (before 8/12/22 7:35:48.000 PM) No Event Sampling ▾

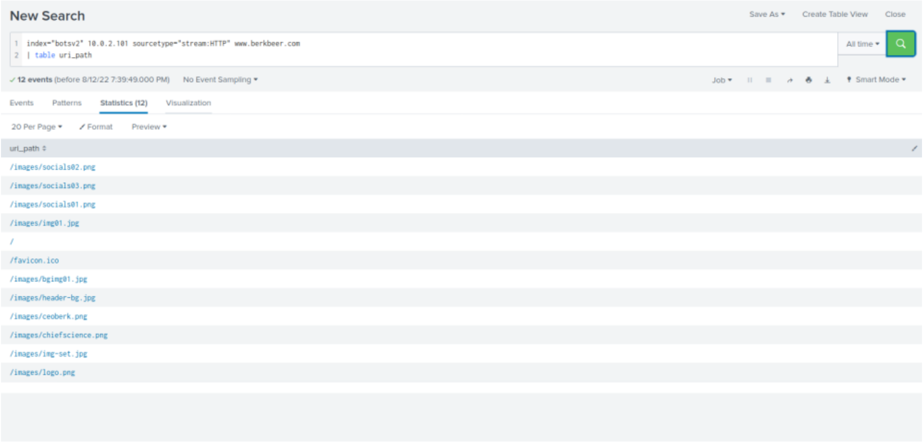
Job ▾ Smart Mode ▾

Events Patterns Statistics (1) Visualization

20 Per Page ▾ Format Preview ▾

site
www.berkbeer.com

Fig.11. Filter further to find what information on the competitor. We add a keyword “BEER” as that’s the industry Amber works for. We can conclude the competitor’s website address is www.berkbeer.com.



New Search

Save As ▾ Create Table View Close

```
1 index="botsv2" 10.0.2.101 sourcetype="stream:HTTP" www.berkbeer.com
2 | table uri_path
```

All time ▾

✓ 12 events (before 8/12/22 7:39:49.000 PM) No Event Sampling ▾

Job ▾ Smart Mode ▾

Events Patterns Statistics (12) Visualization

20 Per Page ▾ Format Preview ▾

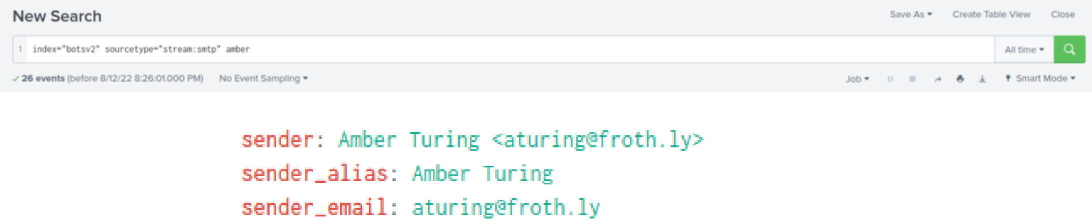
uri_path
/images/socials02.png
/images/socials03.png
/images/socials01.png
/images/img01.jpg
/
/favicon.ico
/images/bgimg01.jpg
/images/header-bg.jpg
/images/ceoberk.png
/images/chiefscience.png
/images/img-set.jpg
/images/logo.png

Fig.12. After finding the competitor’s website, we can narrow Amber’s HTTP traffic to the website and see what resources she accessed. We are looking to know if she obtained the CEO’s name and see if we can find his email address.

/images/ceoberk.png (contains information about the CEO of Berkbeer, and we find their staff email domain is berkbeer.com).

Once we obtain the CEO’s email address, we can build a search query involving SMTP (Simple Mail Transfer Protocol) to view the email traffic between Amber and the CEO.

index="botsv2" sourcetype="stream:smtp"



New Search

Save As ▾ Create Table View Close

```
1 index="botsv2" sourcetype="stream:smtp" amber
```

All time ▾

✓ 26 events (before 8/12/22 8:26:01.000 PM) No Event Sampling ▾

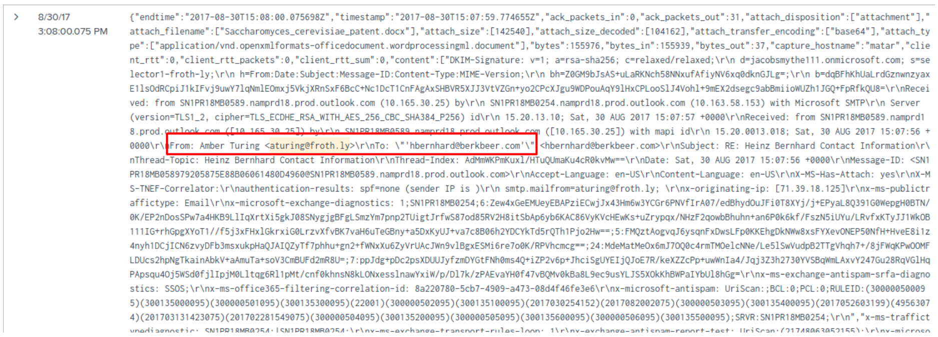
Job ▾ Smart Mode ▾

```
sender: Amber Turing <aturing@froth.ly>
sender_alias: Amber Turing
sender_email: aturing@froth.ly
```

Fig.13. We obtain Amber’s email address aturing@froth.ly and continue the search for communication between Amber and the CEO.

index="botsv2" sourcetype="stream:smtp" aturing@froth.ly berkbeer.com

By viewing the raw data, we can obtain the CEO’s email address: hbernhard@berkbeer.com and his name is Martin Berk.



> 8/30/17
> 3:08:00.075 PM

```
{"endtime":"2017-08-30T15:08:00.075698Z","timestamp":"2017-08-30T15:07:59.774655Z","ack_packets_in":0,"ack_packets_out":31,"attach_disposition":["attachment"],"
attach_filename":["Saccharomyces_cerevisiae_patent.docx"],"attach_size":[142540],"attach_size_decoded":[104162],"attach_transfer_encoding":["base64"],"attach_ty
pe":["application/vnd.openxmlformats-officedocument.wordprocessingml.document"],"bytes":155976,"bytes_in":155939,"bytes_out":37,"capture_hostname":"matar","clie
nt_rtt":0,"client_rtt_packets":0,"client_rtt_sum":0,"content":["DKIM-Signature: v=1; a=rsa-sha256; c=relaxed/relaxed;\r\n d=jacobsmythe111.onmicrosoft.com; s=se
lector1-froth-ly;\r\n h=From:Date:Subject:Message-ID:Content-Type:MIME-Version;\r\n bh=Z0GM9bJsAS+uLaRKNch58NNxufAfiyNV6xq0dknGJLg=;\r\n b=dqBFhKhUaLrdGznwnzyax
E1lsOdRCpiJ1kIFvj9uwY7lqNmlEOmxj5VkjXRnSxF6BcC+Nc1DcT1CnFAgAxSHBVR5XJJ3VtVZGn+yo2CPcXJgu9WDPouAqY9lHxCPLooSlJ4Vohl+9mEX2dsegc9abBmiioWUZh1JGQ+FpRfkQU8=\r\nRecei
ved: from SN1PR18MB0589.namprd18.prod.outlook.com (10.165.30.25) by\r\n SN1PR18MB0254.namprd18.prod.outlook.com (10.163.58.153) with Microsoft SMTP\r\n Server
(version=TLS1_2, cipher=TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA384_P256) id\r\n 15.20.13.10; Sat, 30 AUG 2017 15:07:57 +0000\r\nReceived: from SN1PR18MB0589.namprd1
8.prod.outlook.com ([10.165.30.25]) by\r\n SN1PR18MB0589.namprd18.prod.outlook.com ([10.165.30.25]) with mapi id\r\n 15.20.0013.018; Sat, 30 AUG 2017 15:07:56 +
0000\r\nFrom: Amber Turing <aturing@froth.ly>\r\nTo: \"'hbernhard@berkbeer.com'\" <hbernhard@berkbeer.com>\r\nSubject: RE: Heinz Bernhard Contact Information\r\
nThread-Topic: Heinz Bernhard Contact Information\r\nThread-Index: AdMmWKPmKuxl/HTuQUmaKu4cR0kvMw==\r\nDate: Sat, 30 AUG 2017 15:07:56 +0000\r\nMessage-ID: <SN1
PR18MB058979205875E88B06061480D4960@SN1PR18MB0589.namprd18.prod.outlook.com>\r\nAccept-Language: en-US\r\nContent-Language: en-US\r\nX-MS-Has-Attach: yes\r\nX-M
S-TNEF-Correlator:\r\nauthentication-results: spf=none (sender IP is )\r\n smtp.mailfrom=aturing@froth.ly; \r\nx-originating-ip: [71.39.18.125]\r\nx-ms-publictr
affictype: Email\r\nx-microsoft-exchange-diagnostics: 1;SN1PR18MB0254;6:Zew4xGeEMUeyEBAPziECwjJx43Hm6w3YCGr6PNVfIrA07/edBhydOuJFi0T8XYj/j+EPyaL8Q391G0WepgH0BTN/
0K/EP2nDosSPw7a4HKB9LlIqXrtXi5gkJ08SNygjgBFgLSmzYm7pnp2TUigtJrfwS87od85RV2H8itSbAp6yb6KAC86VyKVcHEwKs+uZrypqx/NHzF2qowbBhuhn+an6P0k6kf/FszN5iUYu/LRvfxKTyJJ1WkOB
111IG+rhGpgXYoT1//f5j3xFHxlGkrxiG0LrzvXfvBK7vaH6uTeGBny+a5DxKyUJ+va7c8B06h2YDCYkTd5rQTh1Pjo2Hw==;5:FMQztAogvqJ6ysqnFxDwsLFp0KKEhgDkNWw8xsFYXevONEP50NfH+HveE8i1z
4nyh1DCjICN6zvyDFb3msxukpHaQJAIQZyTf7phhu+gn2+fWNxXu6ZyVrUAcJWn9vlBgxESMi6re7o0K/RPVhcmcg==;24:MdeMatMeOx6mJ7OQ0c4rmTMOelcNNe/Le5lSwVudpB2TTgVhqh7+/8jFWqKPwOOMF
LDUcs2hpNgTkainAbkV+aAmuTa+soV3CmBUFd2mR8U=;7:ppJdg+pDc2psXDUUJyfzmDYGtFNh0ms4Q+iZP2v6p+JhciSgUYEIjQJoE7R/keXZZcPp+uwWnIa4/Jqj3Z3h2730YVSBqWmLAxvY247Gu28RqVGlHq
PApsqu4Oj5WSd0fjlIpjM0Lltqg6Rl1pMt/cnf0khnsN8kLONxesslnawYxiW/p/Dl7k/zPAEvaYH0f47vBQMv0kBa8L9ec9usYLJS5XOkKhBWPaIYbUl8hGg=\r\nx-ms-exchange-antispam-srfa-diagno
stics: SSOS;\r\nx-ms-office365-filtering-correlation-id: 8a220780-5cb7-4909-a473-08d4f46fe3e6\r\nx-microsoft-antispam: UriScan:;BCL:0;PCL:0;RULEID:(30000050009
5)(300135000095)(300000501095)(300135300095)(22001)(300000502095)(300135100095)(2017030254152)(2017082002075)(300000503095)(300135400095)(2017052603199)(4956307
4)(201703131423075)(201702281549075)(300000504095)(300135200095)(300000505095)(300135600095)(300000506095)(300135500095);SRVR:SN1PR18MB0254;\r\n","x-ms-traffict
ypediagnostic: SN1PR18MB0254:|SN1PR18MB0254;\r\nx-ms-exchange-transport-rules-loop: 1\r\nx-exchange-antispam-report-test: UriScan:(21748063052155);\r\nx-microso
```

Fig.14. Observe an SMTP event’s raw data between Amber and Martin Berk.

We can look at the content of the email between them:

Mr. Bernhard, I was very sorry to hear about the acquisition falling through. I was very excited to work with you in the future. I have to admit, I am a little worried about my future here. I'd love to talk to you about some information I have regarding my work.

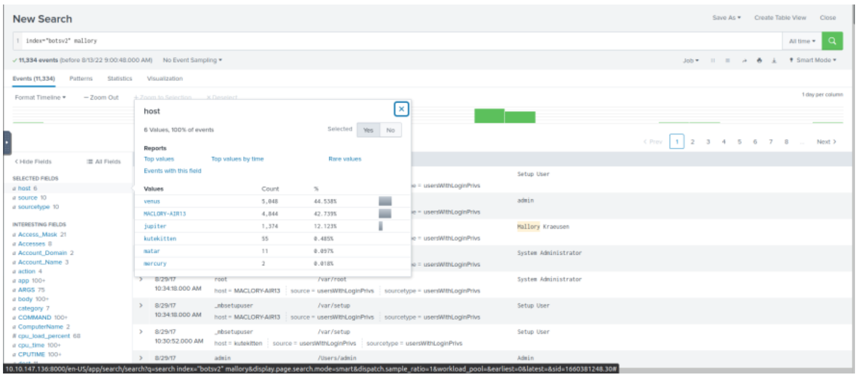
Hello Amber, Great talking with you today, here is my contact information. Do you have a personal email I can reach you at as well? Thank You.

The last event also contains an attachment with patent and sensitive data from the company, which should be confidential and not shared with outsiders.

USB RANSOMWARE ATTACK INVESTIGATION

The scenario involves an individual named Mallory, her Mac OS machine and some encrypted files. The attack occurred due to an infected USB ransomware attack, and she reported that some important PowerPoint files were encrypted and unrecoverable. This investigation aims to find information about the USB used in the attack and information about the malware.

index="botsv2" mallory



```
index="botsv2" mallory
✓ 11,334 events (before 8/13/22 9:00:48.000 AM)   No Event Sampling
```

host
6 Values, 100% of events

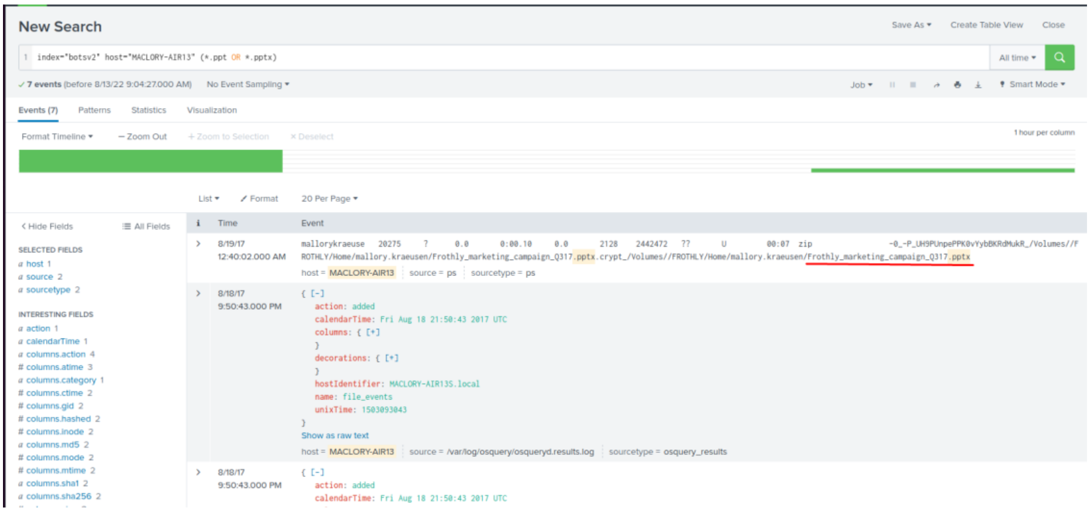
Values	Count	%
venus	5,048	44.538%
MACLORY-AIR13	4,844	42.739%
jupiter	1,374	12.123%
kutekitten	55	0.485%
matar	11	0.097%
mercury	2	0.018%

```
8/29/17 10:34:18.000 AM   root            /var/root    System Administrator
  host = MACLORY-AIR13   source = usersWithLoginPrivs   sourcetype = usersWithLoginPrivs
8/29/17 10:34:18.000 AM   _mbsetupuser    /var/setup   Setup User
  host = MACLORY-AIR13   source = usersWithLoginPrivs   sourcetype = usersWithLoginPrivs
8/29/17 10:30:52.000 AM   _mbsetupuser    /var/setup   Setup User
  host = kutekitten   source = usersWithLoginPrivs   sourcetype = usersWithLoginPrivs
8/29/17                   admin           /Users/admin Admin
```

Fig.15. By using a keyword search for "Mallory" and inspecting the host selected fields, we can obtain her device name "MACLORY-AIR13".

We can now try to locate the corrupted files with the device name. We start to look for a PowerPoint presentation file. Hence we can filter for .ppt or .pptx files.

index="botsv2" host="MACLORY-AIR13" (*.ppt OR *.pptx)

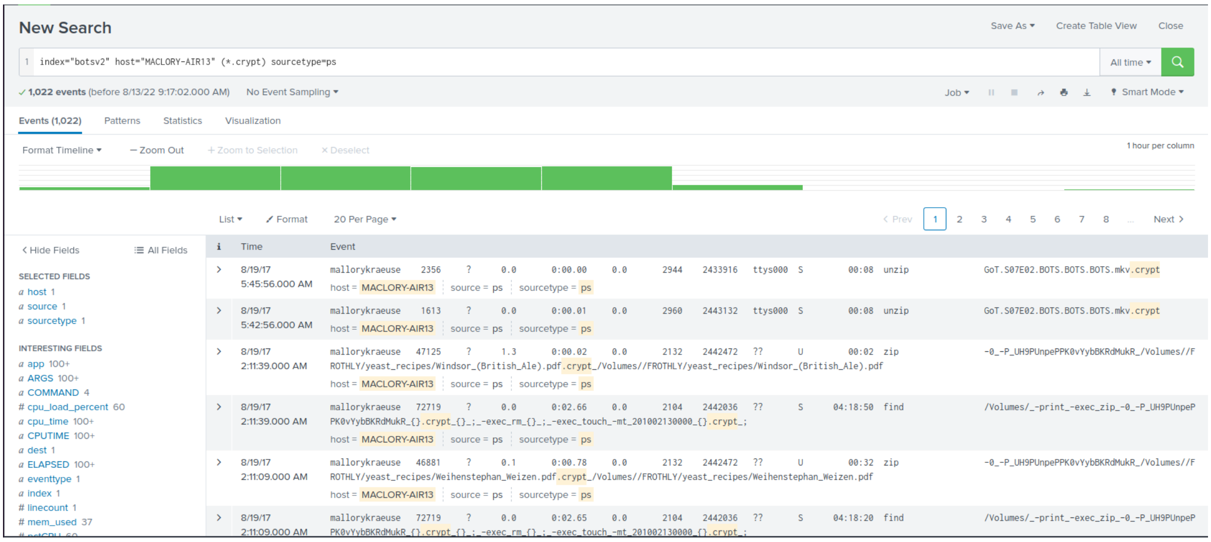


```
index="botsv2" host="MACLORY-AIR13" (*.ppt OR *.pptx)
✓ 7 events (before 8/13/22 9:04:27.000 AM)   No Event Sampling

Time                       Event
8/19/17 12:40:02.000 AM    mallorykraeuse   20275   ?   0.0   0:00.10   0.0   2128   2442472   ??   U   00:07  zip   -0_-P_UH9PUnpePPK0vYybBKRdMukR_/Volumes//FROTHLY/Home/mallory.kraeusen/Frothly_marketing_campaign_Q317.pptx.crypt_/Volumes//FROTHLY/Home/mallory.kraeusen/Frothly_marketing_campaign_Q317.pptx
                           host = MACLORY-AIR13   source = ps   sourcetype = ps
8/18/17 9:50:43.000 PM     { [-]
                             action: added
                             calendarTime: Fri Aug 18 21:50:43 2017 UTC
                             columns: { [+]
                             }
                             decorations: { [+]
                             }
                             hostIdentifier: MACLORY-AIR13S.local
                             name: file_events
                             unixTime: 1503093043
                           }
                           host = MACLORY-AIR13   source = /var/log/osquery/osqueryd.results.log   sourcetype = osquery_results
8/18/17 9:50:43.000 PM     { [-]
                             action: added
                             calendarTime: Fri Aug 18 21:50:43 2017 UTC
```

Fig.16. Quickly sort through and find the encrypted PowerPoint file and file location with only seven events to examine. We can also see that the file has been encrypted using the ".crypt" extension. We can use this information to find other files that have been corrupted.

index="botsv2" host="MACLORY-AIR13" (*.crypt) sourcetype=ps



```
index="botsv2" host="MACLORY-AIR13" (*.crypt) sourcetype=ps
```

✓ 1,022 events (before 8/13/22 9:17:02.000 AM)   No Event Sampling ▾

SELECTED FIELDS
a host 1
a source 1
a sourcetype 1

INTERESTING FIELDS
a app 100+
a ARGS 100+
a COMMAND 4
# cpu_load_percent 60
a cpu_time 100+
a CPUTIME 100+
a dest 1
a ELAPSED 100+
a eventtype 1
a index 1
# linecount 1
# mem_used 37

```
i  Time                  Event
>  8/19/17               mallorykraeuse    2356    ?    0.0    0:00.00    0.0    2944    2433916   ttys000  S       00:08  unzip           GoT.S07E02.BOTS.BOTS.BOTS.mkv.crypt
   5:45:56.000 AM        host = MACLORY-AIR13 | source = ps | sourcetype = ps
>  8/19/17               mallorykraeuse    1613    ?    0.0    0:00.01    0.0    2960    2443132   ttys000  S       00:08  unzip           GoT.S07E02.BOTS.BOTS.BOTS.mkv.crypt
   5:42:56.000 AM        host = MACLORY-AIR13 | source = ps | sourcetype = ps
>  8/19/17               mallorykraeuse   47125    ?    1.3    0:00.02    0.0    2132    2442472   ??       U       00:02  zip             -0_-P_UH9PUnpePPK0vYybBKRdMukR_/Volumes//F
   2:11:39.000 AM        ROTHLY/yeast_recipes/Windsor_(British_Ale).pdf.crypt_/Volumes//FROTHLY/yeast_recipes/Windsor_(British_Ale).pdf
                         host = MACLORY-AIR13 | source = ps | sourcetype = ps
>  8/19/17               mallorykraeuse   72719    ?    0.0    0:02.66    0.0    2104    2442036   ??       S   04:18:50  find            /Volumes/_-print_-exec_zip_-0_-P_UH9PUnpeP
   2:11:39.000 AM        PK0vYybBKRdMukR_{}.crypt_{}_;_-exec_rm_{}_;_-exec_touch_-mt_201002130000_{}.crypt_;
                         host = MACLORY-AIR13 | source = ps | sourcetype = ps
>  8/19/17               mallorykraeuse   46881    ?    0.1    0:00.78    0.0    2132    2442472   ??       U       00:32  zip             -0_-P_UH9PUnpePPK0vYybBKRdMukR_/Volumes//F
   2:11:09.000 AM        ROTHLY/yeast_recipes/Weihenstephan_Weizen.pdf.crypt_/Volumes//FROTHLY/yeast_recipes/Weihenstephan_Weizen.pdf
                         host = MACLORY-AIR13 | source = ps | sourcetype = ps
>  8/19/17               mallorykraeuse   72719    ?    0.0    0:02.65    0.0    2104    2442036   ??       S   04:18:20  find            /Volumes/_-print_-exec_zip_-0_-P_UH9PUnpeP
   2:11:09.000 AM        PK0vYybBKRdMukR_{}.crypt_{}_;_-exec_rm_{}_;_-exec_touch_-mt_201002130000_{}.crypt_;
```

Fig.17. The result shows that 1022 files were infected by this ransomware attack on Mallory's device.

Now, we move on to identifying the USB device used, using the information given. We know the perpetrator is an employee named Kevin Lagerfield and that the USB drive was connected to Mallory's personal MacBook kutekitten. When Mallory ran the malware, it obfuscated itself during the execution.

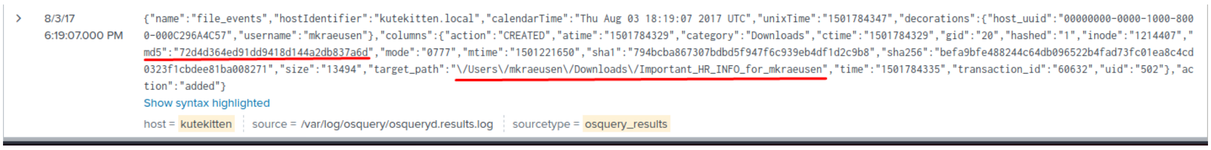
There are two options to find information about the USB: looking through the registry logs in the Windows Event Logs for the event of a USB being connected to the device and using the estimated time of infection or finding the ransomware and then finding information on how the ransomware was introduced.

To find the malware, we will be using OSquery. OSquery exposes an operating system as a high-performance relational database. This allows you to write SQL queries to explore operating system data. With OSquery, SQL tables represent abstract concepts such as running processes, loaded kernel modules, open network connections, browser plugins, hardware events or file hashes.

I started with Mallory's Users folder to look for the malware, using OSquery syntax.

After examining the library, documents and download folder for suspicious files, we located a file named "Important_HR_INFO_for_mkraeusen", which I believe to be the ransomware.

index="botsv2" host=kutekitten sourcetype=osquery_results "columns.target_path"="/Users/mkraeusen/Downloads/Important_HR_INFO_for_mkraeusen"



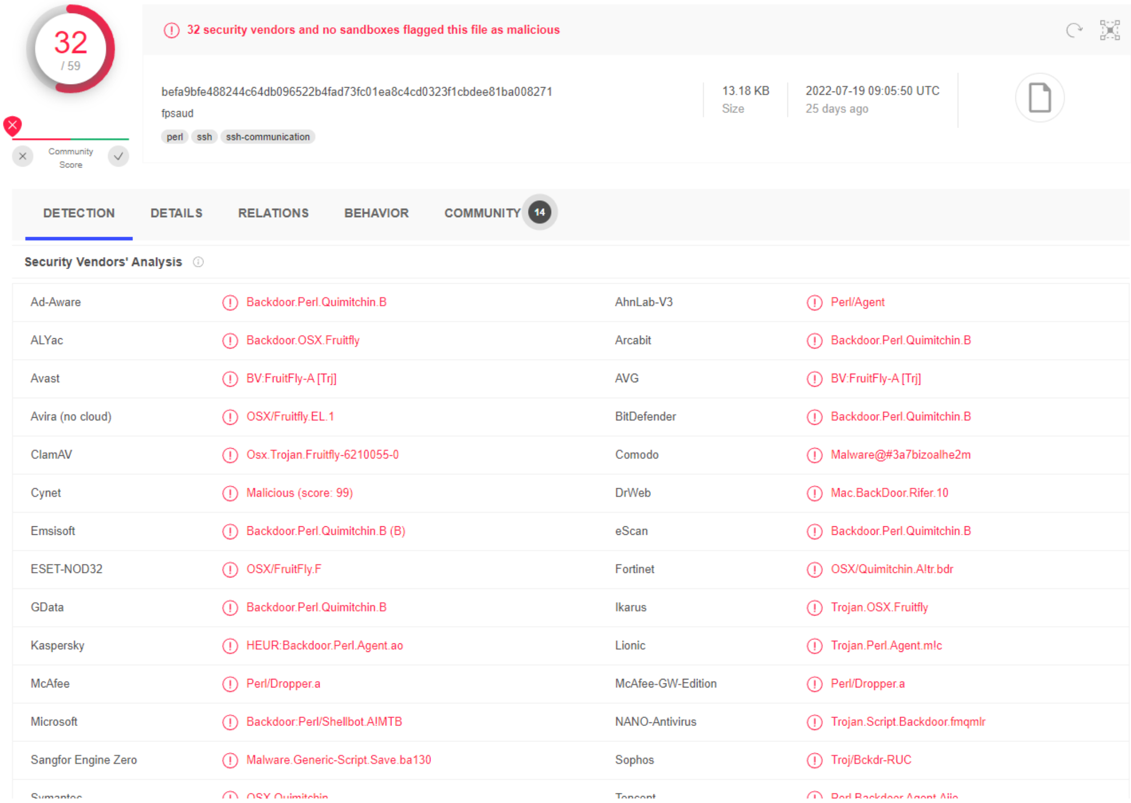
8/3/17
6:19:07.000 PM

{"name":"file_events","hostIdentifier":"kutekitten.local","calendarTime":"Thu Aug 03 18:19:07 2017 UTC","unixTime":"1501784347","decorations":{"host_uuid":"00000000-0000-1000-8000-000C296A4C57","username":"mkraeusen"},"columns":{"action":"CREATED","atime":"1501784329","category":"Downloads","ctime":"1501784329","gid":"20","hashed":"1","inode":"1214407","md5":"72d4d364ed91dd9418d144a2db837a6d","mode":"0777","mtime":"1501221650","sha1":"794bcba867307bdbd5f947f6c939eb4df1d2c9b8","sha256":"befa9bfe488244c64db096522b4fad73fc01ea8c4cd0323f1cbdee81ba008271","size":"13494","target_path":"\/Users\/mkraeusen\/Downloads\/Important_HR_INFO_for_mkraeusen","time":"1501784335","transaction_id":"60632","uid":"502"},"action":"added"}

host = kutekitten | source = /var/log/osquery/osqueryd.results.log | sourcetype = osquery_results

Fig.18. Details of the suspicious file download by "mkraeusen".

We can carry out a search in Virus total with the md5 of the file to confirm if the file is ransomware.



32 / 59

Community Score

32 security vendors and no sandboxes flagged this file as malicious

befa9bfe488244c64db096522b4fad73fc01ea8c4cd0323f1cbdee81ba008271
fpsaud
perl ssh ssh-communication

13.18 KB
Size

2022-07-19 09:05:50 UTC
25 days ago

DETECTION DETAILS RELATIONS BEHAVIOR COMMUNITY 14

Security Vendors' Analysis

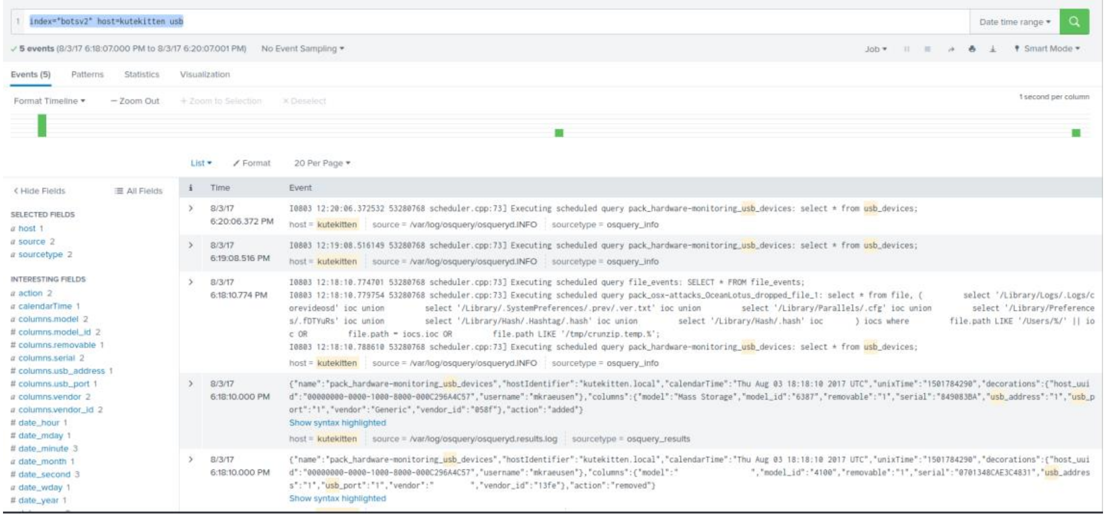
Ad-Aware	Backdoor.Perl.Quimitchin.B	AhnLab-V3	Perl/Agent
ALYac	Backdoor.OSX.Fruitfly	Arcabit	Backdoor.Perl.Quimitchin.B
Avast	BV:FruitFly-A [Trj]	AVG	BV:FruitFly-A [Trj]
Avira (no cloud)	OSX/Fruitfly.EL.1	BitDefender	Backdoor.Perl.Quimitchin.B
ClamAV	Osx.Trojan.Fruitfly-6210055-0	Comodo	Malware@#3a7bizoalhe2m
Cynet	Malicious (score: 99)	DrWeb	Mac.BackDoor.Rifer.10
Emsisoft	Backdoor.Perl.Quimitchin.B (B)	eScan	Backdoor.Perl.Quimitchin.B
ESET-NOD32	OSX/FruitFly.F	Fortinet	OSX/Quimitchin.A!tr.bdr
GData	Backdoor.Perl.Quimitchin.B	Ikarus	Trojan.OSX.Fruitfly
Kaspersky	HEUR:Backdoor.Perl.Agent.ao	Lionic	Trojan.Perl.Agent.m!c
McAfee	Perl/Dropper.a	McAfee-GW-Edition	Perl/Dropper.a
Microsoft	Backdoor:Perl/Shellbot.A!MTB	NANO-Antivirus	Trojan.Script.Backdoor.fmqmlr
Sangfor Engine Zero	Malware.Generic-Script.Save.ba130	Sophos	Troj/Bckdr-RUC
Symantec	OSX.Quimitchin	Tencent	Perl.Backdoor.Agent.Alio

Fig.19. Virustotal search of the ransomware details through the md5 value.

Thirty-two vendors reported this as malware, so we consider it the malware Mallory executed on 8/3/17 at 6:19 pm. We can deduce that the USB was plugged before the malware was run; hence we can set a new filter to display events before the malware execution.

index="botsv2" host=kutekitten USB

Display USB events within a time range of sixty seconds before the execution of malware.



```
index="botsv2" host=kutekitten usb
```

✓ 5 events (8/3/17 6:18:07.000 PM to 8/3/17 6:20:07.001 PM) No Event Sampling ▾

SELECTED FIELDS
a host 1
a source 2
a sourcetype 2

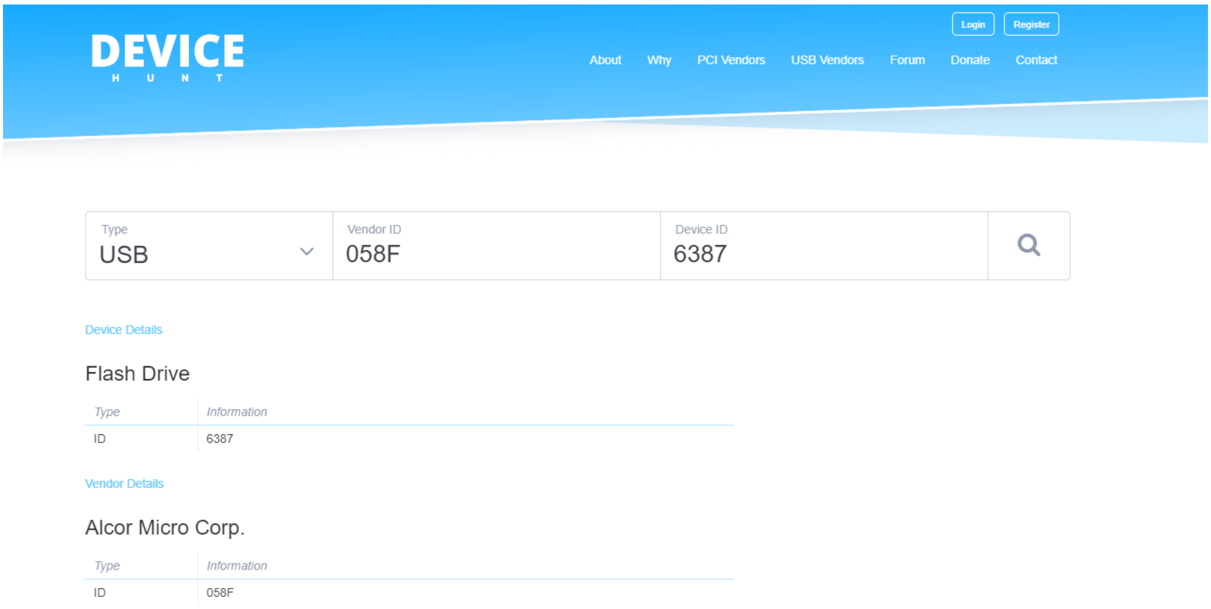
INTERESTING FIELDS
a action 2
a calendarTime 1
a columns.model 2
# columns.model_id 2
# columns.removable 1
a columns.serial 2
# columns.usb_address 1
# columns.usb_port 1
a columns.vendor 2
a columns.vendor_id 2
# date_hour 1
# date_mday 1
# date_minute 3
a date_month 1
# date_second 3
a date_wday 1
# date_year 1

i	Time	Event
>	8/3/17 6:20:06.372 PM	I0803 12:20:06.372532 53280768 scheduler.cpp:73] Executing scheduled query pack_hardware-monitoring_usb_devices: select * from usb_devices; host = kutekitten source = /var/log/osquery/osqueryd.INFO sourcetype = osquery_info
>	8/3/17 6:19:08.516 PM	I0803 12:19:08.516149 53280768 scheduler.cpp:73] Executing scheduled query pack_hardware-monitoring_usb_devices: select * from usb_devices; host = kutekitten source = /var/log/osquery/osqueryd.INFO sourcetype = osquery_info
>	8/3/17 6:18:10.774 PM	I0803 12:18:10.774701 53280768 scheduler.cpp:73] Executing scheduled query file_events: SELECT * FROM file_events; I0803 12:18:10.779754 53280768 scheduler.cpp:73] Executing scheduled query pack_osx-attacks_OceanLotus_dropped_file_1: select * from file, (select '/Library/Logs/.Logs/corevideosd' ioc union select '/Library/.SystemPreferences/.prev/.ver.txt' ioc union select '/Library/Parallels/.cfg' ioc union select '/Library/Preferences/.fDTYuRs' ioc union select '/Library/Hash/.Hashtag/.hash' ioc union select '/Library/Hash/.hash' ioc) iocs where file.path LIKE '/Users/%/' || ioc OR file.path = iocs.ioc OR file.path LIKE '/tmp/crunzip.temp.%'; I0803 12:18:10.788610 53280768 scheduler.cpp:73] Executing scheduled query pack_hardware-monitoring_usb_devices: select * from usb_devices; host = kutekitten source = /var/log/osquery/osqueryd.INFO sourcetype = osquery_info

>	8/3/17 6:18:10.000 PM	{"name":"pack_hardware-monitoring_usb_devices","hostIdentifier":"kutekitten.local","calendarTime":"Thu Aug 03 18:18:10 2017 UTC","unixTime":"1501784290","decorations":{"host_uuid":"00000000-0000-1000-8000-000C296A4C57","username":"mkraeusen"},"columns":{"model":"Mass Storage","model_id":"6387","removable":"1","serial":"849083BA","usb_address":"1","usb_port":"1","vendor":"Generic","vendor_id":"058f"},"action":"added"}
		Show syntax highlighted
		host = kutekitten source = /var/log/osquery/osqueryd.results.log sourcetype = osquery_results

>	8/3/17 6:18:10.000 PM	{"name":"pack_hardware-monitoring_usb_devices","hostIdentifier":"kutekitten.local","calendarTime":"Thu Aug 03 18:18:10 2017 UTC","unixTime":"1501784290","decorations":{"host_uuid":"00000000-0000-1000-8000-000C296A4C57","username":"mkraeusen"},"columns":{"model":" ","model_id":"4100","removable":"1","serial":"0701348CAE3C4831","usb_address":"1","usb_port":"1","vendor":" ","vendor_id":"13fe"},"action":"removed"} Show syntax highlighted

Fig.20. Five events are displayed, and the last two are relevant as they portray the addition and removal of a USB device. We can obtain the model id and vendor id of the USB: model_id": "6387”, vendor id: “058f”.



Type
USB

Vendor ID
058F

Device ID
6387

Device Details

Flash Drive

Type	Information
ID	6387

Vendor Details

Alcor Micro Corp.

Type	Information
ID	058F

Fig.21. Device Hunt website searches for more information on the USB and the USB vendor.