University of Piraeus

School of Information and Communication Technologies

Department of Digital Systems

Postgraduate Program of Studies

MSc Digital Systems Security

Master's Thesis

CTI sharing optimizations and automating threat detection based on actionable intelligence

Supervisor Professor: Christos Xenakis

Name-Surname	E-mail	Student ID.
Marios Karatisoglou	m.karatisoglou@ssl-unipi.gr	MTE2012

Piraeus

05/07/2022

Abstract

Security for businesses and organizations is essential to protect operational activities, trust relationships with potential clients and financial viability. Increased interest for research concerning cybersecurity issues has been shown recently, while at the same time professionals of this sector are employed to ensure safety. In turn, the efficacy and performance of both the researchers and professionals rely on the information provided by Cyber Threat Intelligence (CTI) infrastructures. Automation of procedures regarding the collection, harmonization and processing of information is of utmost importance for CTI, in order to effectively relay to the community intelligence concerning newly emerged threats. Nevertheless, the process regarding the transfer of knowledge between CTI and cybersecurity specialists is based on frameworks and procedures that are not in line with the needs and standards of modern times, being performed through obsolete methods and manual labor. In this paper, we propose BRIDGE, the first tool that streamlines the flow of intelligence between CTI and cybersecurity professionals, by taking advantage of the STIX standard, utilizing blockchain technology and automatically converting the intelligence needed in the form that researchers and other professionals require. Our experimental results demonstrate the efficiency of BRIDGE in terms of swiftness and performance improvement compared to the mainstream approach.

Table of Content

Acknowledgements

First and foremost, I would like to thank my supervisor, Aristeidis Farao. He has been an ideal teacher, mentor, and thesis supervisor, offering his advice and guidance willingly every time it was necessary for the progress of this thesis. I'm very grateful for my time working with Aristeidis.

My sincere thanks also goes to professor Christos Xenakis. He was the one who offered me the opportunity to work on the field of Cyber Threat Intelligence and I cannot thank him enough for introducing me to this field of cybersecurity. Moreover, I gratefully recognize the knowledge he shared with us during his lectures, which helped me build the background needed to work on this thesis.

Last of all, I am grateful for my parents who supported me through my studies. This thesis would have been impossible if it was not for their aid.

Introduction

Threat intelligence is rapidly becoming a business priority. This happens due to the increasing sophistication and scale of cyber attacks. Teams or individuals that perform cyber crime demonstrate an improvement in their tactics, techniques and procedures (TTPs), making it very difficult and challenging to investigate their activity [1]. A remarkable example of that is the WannaCry ransomware attack. Within a year, it has managed to spread over 150 countries, infecting not less than 230,000 computers [2]. If threat data analysis is performed correctly, the products of threat intelligence can be genuinely useful to a business, helping at the detection and prevention of large scale cyber attacks. Having a cyber threat intelligence (CTI) program as part of the organization, it provides the ability to act proactively in current threats.

A quick review in the history of threat intelligence would be helpful in understanding the concepts around cyber threat intelligence. The idea of intelligence has originated from the military. Armies in times of war gathered all the information possible, not only from the environment but also from the opponent army. Understanding the opponent's tactics and techniques gives them the upper hand in the field of battle. According to [3], "Military Intelligence is a military discipline that focuses on the gathering, analysis, protection and dissemination of information of both strategic and tactical value". This information should be used when making decisions and the actions that an army will take should be based on that information. This includes information about the enemy, terrain and weather in an area of operations or area of interest [3]. But how could this information about the enemy be gathered? The most ancient method is spying. Spying is one of the most important tools in information gathering. The Romans had a network of spies and embassies that they used to collect valuable information, including the environment and socio-political information about neighboring states and people [3]. There are two types of intelligence, strategic and tactical and the only difference between them is the

scope of appliance. Strategic intelligence is used to formulate long-term policies on the national and international scale and is concerned with broad issues such as economics, military capabilities of foreign countries and political assessments [3]. On the other hand, tactical intelligence is more focused on the specific objectives and situation of military commanders in the field.

Since then, those types of intelligence have evolved. As time passed by, wars and military scopes were far from where they had started. With the technology, tools and artifacts continuously changing, so did war. In the 90's when the internet was introduced, military activity started to move to the cyber realm. This is when the Cyber Threat Intelligence was born, when the field of battle has shifted over computer networks.

Taking the previous into the field of cyber security, CTI should apply all those characteristics of threat intelligence against cyber crime. According to [4], Cyber Threat Intelligence is defined as evidence-based knowledge, including context, mechanisms, indicators, implications and actionable advice, about an existing or emerging menace or hazard to assets that can be used to inform decisions regarding the subject's response to that menace or hazard.

In order for an organization to properly implement a CTI program, it is very important first to understand its definition. There are also some concepts that form the basis of understanding CTI, cyber threat and cyber attack. The definition of cyber threat in the Oxford English Dictionary [5] is "the possibility of malicious attempts to damage or disrupt a computer network or system". It is also stated in [5] that cyber attack is "an attempt by hackers to damage or destroy a computer network or system".

Overall the contribution of this work are the following:

- implement a decentralized CTI sharing platform based on blockchain
- generate Sigma rules based on actionable intelligence in CTI reports
- SOCs can automatically generate queries for their SIEM based on indicators provided by CTI reports
- Cyber Threat Intelligence lifecycle has progressed significantly by filling the gap between CTI and Security Operations Centers.

Cyber Threat Intelligence Lifecycle

Cyber threat intelligence is implemented by following a specific lifecycle. According to the cybersecurity technology company Crowdstrike [6], the intelligence cycle consists of six stages, requirements, collection, processing, analysis, dissemination and feedback. This flow ensures that the CTI activities are aligned with the organization's objectives and provide actionable information with the required meanings. Following this lifecycle, an organization can achieve constant improvement, which is one of the most important aspects in order to keep the cyber threat intelligence productive and effective.

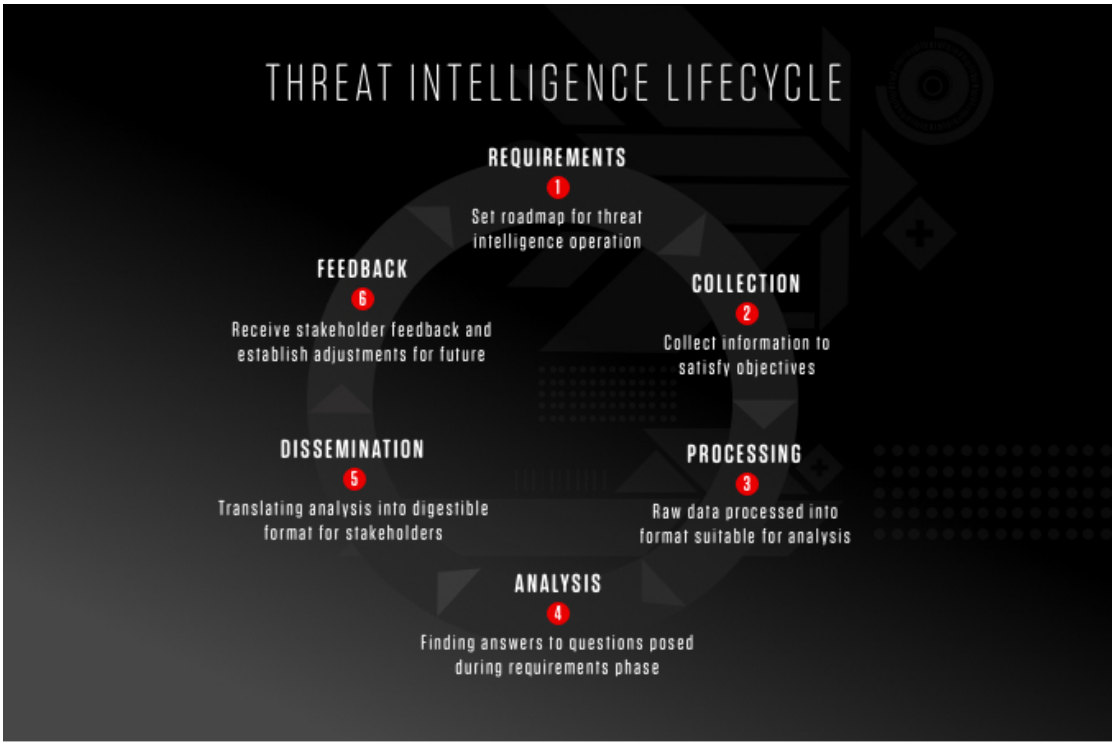

Figure 1. Cyber Threat Intelligence lifecycle [6]

Requirements

This is related to the direction of all of its activities. In this stage, the goals and purpose of the cyber threat intelligence program are defined. It should also be clearly

defined what are the information assets and business processes that need to be protected, alongside with the potential impacts of losing those assets or interrupting those processes. Those assets should also be prioritized according to what is more important to protect. It is also crucial to define who the possible attackers are and what could be their motivations. By knowing that, the next thing to decide is what specific actions should be taken to strengthen their defenses against a future attack.

Collection

Once the requirements in the previous stage are defined, the team must seek to collect the information required to achieve those objectives. The team will usually search for traffic logs from internal networks or devices. Another method is scanning publicly available data sources and scraping relevant forums, social media and websites. Rich source of intelligence can be a subscription to threat data feeds from industry organizations or cybersecurity vendors. This process of gathering information to address the intelligence requirements is called collection.

Processing

By processing it means fitting raw data in a suitable format for further analysis. Different collection methods often require different means of processing, whether by humans or machines running specific algorithms.

Let's take as an example a case where many logs have been collected from a firewall device. By processing those logs, the team could identify IP addresses that are indicators of compromise (IoC). For example, an IP address that tried to access multiple ports could be an indicator of port scanning.

Analysis

Once a dataset has been created by processing the raw data collected in the second stage, the team must then conduct a thorough analysis to find answers to the questions posed in the requirements phase. The product of this stage are reports that

summarize the data for decision makers. Those reports must be concise, avoid confusing and overly technical terms, articulate the issues in business terms and include a recommended course of action [7]. It is important for a report to be clear, in order to communicate better with non-technical leaders.

Dissemination

Dissemination is getting the intel to the targeted audience. It is important that this is made through secure channels. Most cybersecurity organizations have at least six teams that can benefit from threat intelligence and how the analysis is presented depends on the audience [7]. In order to assemble a useful report, the CTI teams must know for each audience what threat intelligence they need, how should the intelligence be presented and how often new updates should be provided.

Feedback

It is very critical for the CTI team to get feedback from the audiences that exchange threat intelligence. An accurate and constructive feedback will assist in boosting the outcome of the next generation of lifecycle. A feedback that would be effective for the improvement of the CTI life cycle would be what types of data to collect, how to process and enrich the data to make useful information, how to analyse and present it as actionable intelligence and how it should be disseminated to each audience [7].

CTI Role in Security Operations Center

Cyber threat intelligence provides an antidote to many of Security Operations Center's (SOC) problems. A SOC is responsible for monitoring a lot of information related to the organization's infrastructure. Specifically, they monitor for potential threats in order to know when something malicious is happening, if any actions should be taken and when the Cyber Incident Response Team (CSIRT) should be alerted. This is done by configuring security sensors like firewalls, IDS and IPS. By doing that in a timely manner, the team will be able to detect suspicious network activity and contain active threats, which can be remediated using available technologies.

Usually, the amount of alerts generated by the network that is being monitored are overwhelming. Alert fatigue makes analysts take alerts less seriously than they should, while many are never investigated at all. When asked to identify their top incident response challenges, 36% of the cybersecurity professionals surveyed said, "keeping up with the volume of security alerts" [8]. Forty-two percent of cybersecurity professionals say that their organization ignores a significant number of security alerts because they can't keep up with the volume. When asked to estimate the percentage of security alerts ignored at their organization, 34% say between 26% and 50%, 20% of cybersecurity professionals say their organization ignores between 50% and 75% of security alerts, and 11% say their organization ignores more than 75% of security alerts. This is a very high amount of alerts not being investigated. In its 2018 State of the SOC report [9], SIEM provider Exabeam revealed that SOCs are understaffed according to 45 percent of professionals who work in them, and of those, 63 percent think they could use anywhere from two to 10 additional employees. Cisco's 2018 Security Capabilities Benchmark study [10] found that organizations can investigate only 56 percent of the security alerts they receive on a given day, and of those investigated alerts, only 34 percent are deemed legitimate.

A survey conducted by Recorded Future [7] indicates that most SOC teams can only investigate 56% of the alerts they receive, and only 34% of investigated

alerts are deemed legitimate. As much as 25% of a security analyst's time is spent investigating false positives, meaning for every hour an analyst works, they waste 15 minutes chasing false positives instead of addressing real threats that put their organization at risk. In 2019, this has led five times as many SOC analysts to believe their primary job responsibility is to "reduce the time it takes to investigate alerts." That approach is taking a serious toll on an already overworked and understaffed workforce. Eight out of 10 security teams report that their SOC has experienced at least 10% and up to more than 50% analyst churn in the past year.

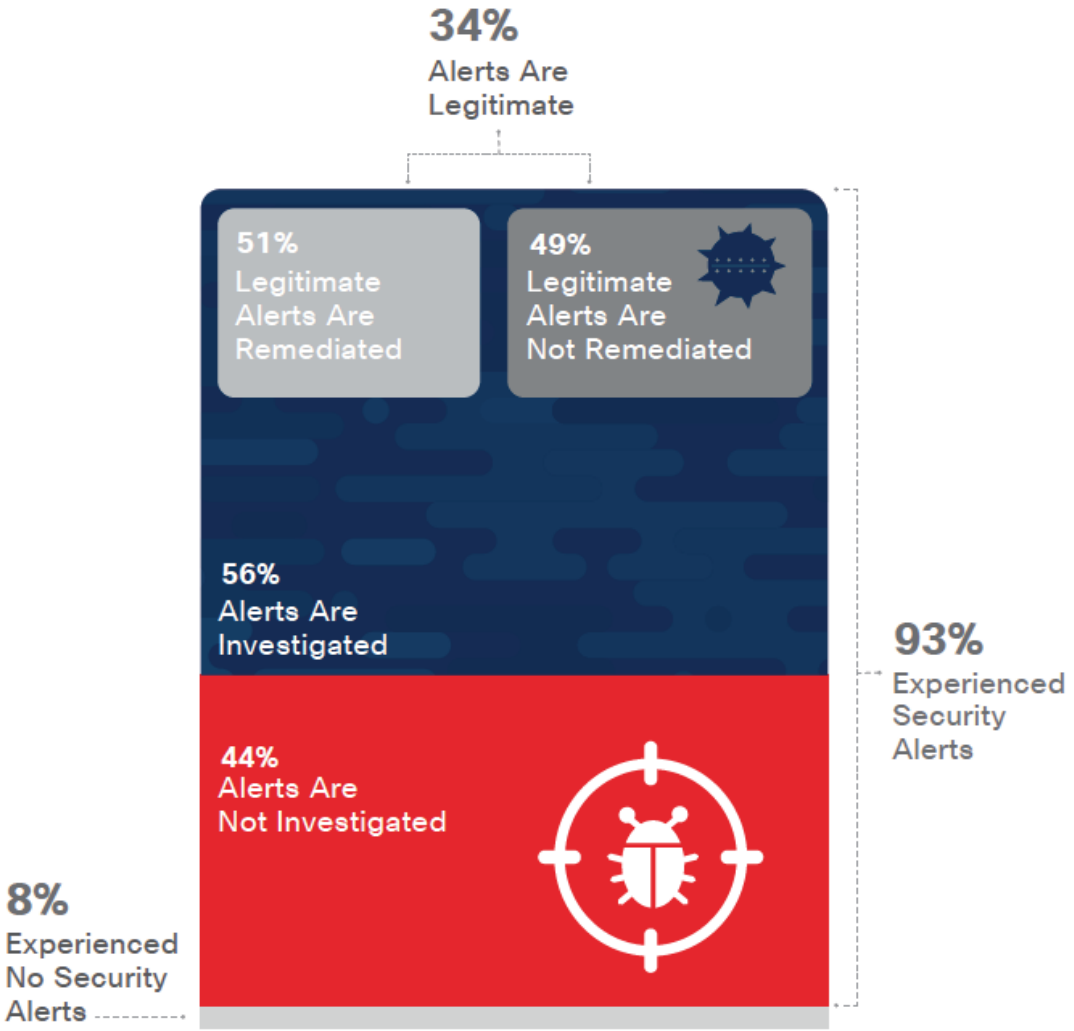

Figure 2. Total alerts and the amount that is being investigated [7].

Context is a very important matter for the SOC unit. Threat Intelligence for the SOC is about enriching internal alerts with the external information and context

harvested by the CTI team and helping to make risk-based decisions. Context is not only critical for rapid triage, but also for scoping and containing incidents. The best way to eliminate the challenges that SOC analysts face is by enriching the information they are receiving, so they are not overwhelmed by alerts that end up being false positives. From the total amount of false positive alerts that are generated, a great part of them could have been detected by an automated system, with no human interaction.

For the rest of the alerts left, the context provider will allow the SOC analyst to perform rapid triage, since most alerts will come with the right context and the decision of either escalating or discarding it will be less time consuming. It will also enable them to have a better scope of the threat, since the information of how far the threat can spread will already be there. As a result, this knowledge will help in better incident containment.

CTI Issues and Challenges

Threat data quality

Threat intelligence sharing is the biggest part of cooperative and collaborative cybersecurity. There are many sources and open communities that provide threat intelligence in order to help organizations and individuals collect actionable information. However, on many occasions the quality of information received by sources varies significantly. An example of this is information that is mistakenly categorized as Indicators of Compromise (IoC). In order to be able to draw conclusions about the quality of information it is necessary to monitor the data provided by the source continuously.

The research in [11] is focused on this issue. In order to be able to quantify the quality of a threat intelligence source, it defines some characteristics that should be present in its information. Those characteristics are extensiveness, maintenance, false positives, verifiability, intelligence, interoperability, compliance, similarity, timeliness and completeness. Based on those parameters, the final goal is to derive a trust indicator to assess the trust in the quality of each CTI source.

Threat Data Overload

The number of threat data sources has increased drastically. A CTI practitioner must select carefully the sources from which the data will be consumed. After the data is collected, an organization will be facing an enormous volume of information. In addition, a lack of staff expertise can make the process of creating actionable intelligence even more difficult. The survey at [12] indicates this issue, with 70% of the participants stating that threat intelligence is too voluminous and/or complex to provide actionable intelligence.

Human brain is not capable of coping with large amounts of data. The best practice to deal with voluminous information is automation tools. This is supported by the research in [13], where 80% of the respondents agree that deploying a threat intelligence platform can help the organization to automate threat intelligence. The

research in [14] presents such a tool to gather cyber threat information from online sources automatically. Given a number of sources like websites as an input, it first uses Support Vector Machine (SVM) machine learning algorithm to classify articles in two categories, those who contain IoCs and those who do not. It then parses the plain text and uses regular expressions to find possible IoCs. In order to reduce the amount of false positives, it uses the Natural Language Processing (NLP) algorithm to extract the grammar terms in a sentence. Finally, a dependency graph is created based on those terms to describe the grammar structure and decide with more confidence if a possible IoC is true or false positive.

CTI Sharing

There are many problems in CTI sharing and they are not limited only to technical issues. There are privacy and legal issues that apply to CTI sharing. An organization should be careful not to expose any personal identifiable information when sharing threat data. The General Data Protection Regulations (GDPR) legal rules could be violated when sharing personal identifiable information without the owner's consent, leading to financial penalties and loss of revenue.

CTI shared with malicious groups disguised as legitimate CTI consumers could have detrimental impact on the organization. Attackers that get their hands on an organization's threat information could expose that knowledge. This will lead to reputation damage and loss of revenue. The research in [15] demonstrates an innovative way to protect the data confidentiality using the blockchain technology. They achieved a successful partitioning of the network with channels, using the Hyperledger technology. This protects participants from sharing highly-sensitive data with unintended parties.

An additional challenge that currently exists in CTI sharing is that there is no consistency in the use of standards that are used to describe cyber threat intelligence. Currently, there are a number of standards that support information sharing and automated cyber security. Such a standard is CybOx (Cyber Observable expression XML) [16]. Another popular standard is STIX (Structured Threat Information expression) [17]. STIX uses CybOx to describe cyber threat info so it can be shared,

stored and analyzed. A CTI producer is provided with a handful of standards to use when documenting and then sharing its CTI data. On the other hand, a CTI consumer will always choose to consume CTI from multiple sources that will not always use the same CTI standard. As a result, the organizations must perform a significant amount of effort to manage all that information received in different standards format.

Blockchain Technology

Blockchain technology has been in the spotlight for the last few years, with Bitcoin and Ethereum [18]. Such virtual currencies that are built on blockchain technology are a decentralized data recording system. A blockchain technology implements a distributed database [19]. This database consists of a continuously growing list of data records,which are also known as ‘ledgers’. Each new record appended to the ledger is validated by the participants in the network. If a consensus is reached, the new record is added to the ledger [20]. In order to identify valid participants in the network, blockchain utilizes a peer-to-peer architecture with public-private key cryptographic mechanisms. The cryptographic security of the blockchain is its main advantage, since it is almost impossible to change a block written to the chain [18].

Blockchain can be viewed as a shared record among participating parties. Each party has a copy of that record. When a new input needs to be written to the record, it is verified by all parties, which will now see the same updated version of the record. That new input is called a transaction. Transactions performed in blockchain technology are written into blocks in a chain. The chain grows continuously by adding a chain for each transaction that occurs.

One significant feature of a blockchain is decentralization. The blockchain can operate in a decentralized environment by integrating many key technologies such as cryptographic hash, digital signature (based on asymmetric cryptography), and distributed consensus mechanism. When a transaction is taking place, it is approved and published in a decentralized way.

Chain Structure

In a blockchain, all blocks have a parent except for the first one. Each block also holds the address of the parent block. That data structure is similar to a single linked list. A visualization of the blockchain structure is depicted in Figure 3.

Figure 3. Representation of a blockchain [18]

For each block, its hash value is calculated using the stored data in the block. This sets a strong validation mechanism for not changing the data in the block, since the hash should also be changed if the data in the block is changed. This way the integrity principle is achieved when storing data in a blockchain. To form the structure of a single linked list, the hash of each block is held by its following block.

A block consists of two parts, the block header and the block body. The block header contains the following fields:

- Version: Describes the structure of the data inside the block. This is used so that computers can read the contents of each block correctly.
- Hash of parent block: A 256-bit hash that points to the previous / parent block.
- Merkle Root: All of the transactions inside the block hashed together to form a single line of text.
- Timestamp: The current time
- Nonce: A 4-byte field that usually starts with 0 and shows increments for each hash.
- target: Calculated from the difficulty, which is a value set by the bitcoin network to regulate how difficult it is to add a block of transactions to the blockchain.

The greater the difficulty, the lower the target, and the more difficult it is to find a block hash that is below this value. The difficulty is a number that regulates how long it takes for miners to add new blocks of transactions to the blockchain. This number is set according to the network's requirements.

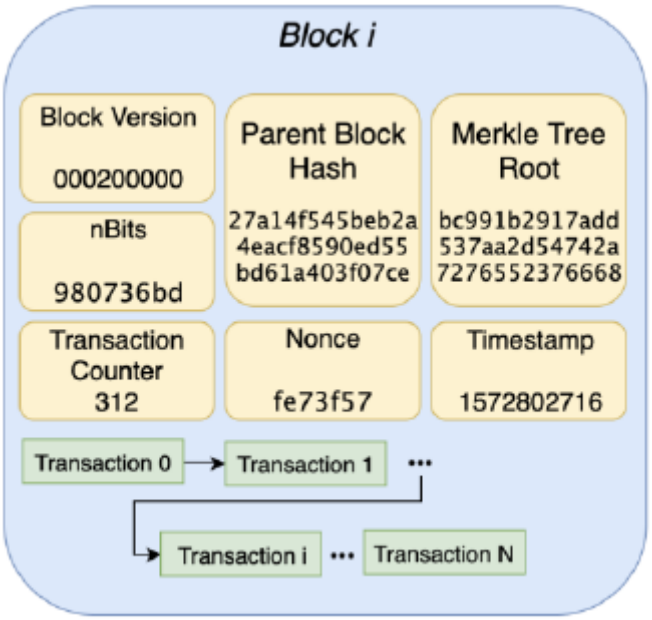

Figure 4. Block Structure

The hash does not derive only from the data of the block. The block data is hashed with an extra number which is called nonce. Nonce is an arbitrary number used only once in a cryptographic communication. Different nonce values are used to get a block hash below the target value. Once a nonce that works has been found, the block is "solved" and all of the transactions in this block are added to the blockchain.

For the blockchain system to work, the entered block data must be validated by all miners (nodes). Consistent reconciliation between all nodes is required. Reconciliation between miners is called Consensus. How Consensus will be reached is the main problem that must be solved for each blockchain system. This is the disadvantage of a decentralized system, where there is no central node that ensures that the ledger in the distributed nodes is the same. Two of the most common protocols that achieve consensus are Proof of Work (PoW) and Proof of Stake (PoS) [18]. Bitcoin uses the PoW protocol to achieve Consensus among the blockchain network nodes.

Smart Contracts

A smart contract (SC) is a computer protocol intended to digitally facilitate, verify, or enforce the negotiation or performance of a contract. [18]. It is stored on a blockchain and runs only when predetermined conditions are met. Smart contracts allow the performance of credible transactions without third parties. They can also automate a workflow, triggering the next action when conditions are met. The main purpose of smart contracts is to simplify transactions between parties by removing intermediate steps involved in business processes. They aim to reduce delays, risks of errors, authenticity and credibility.

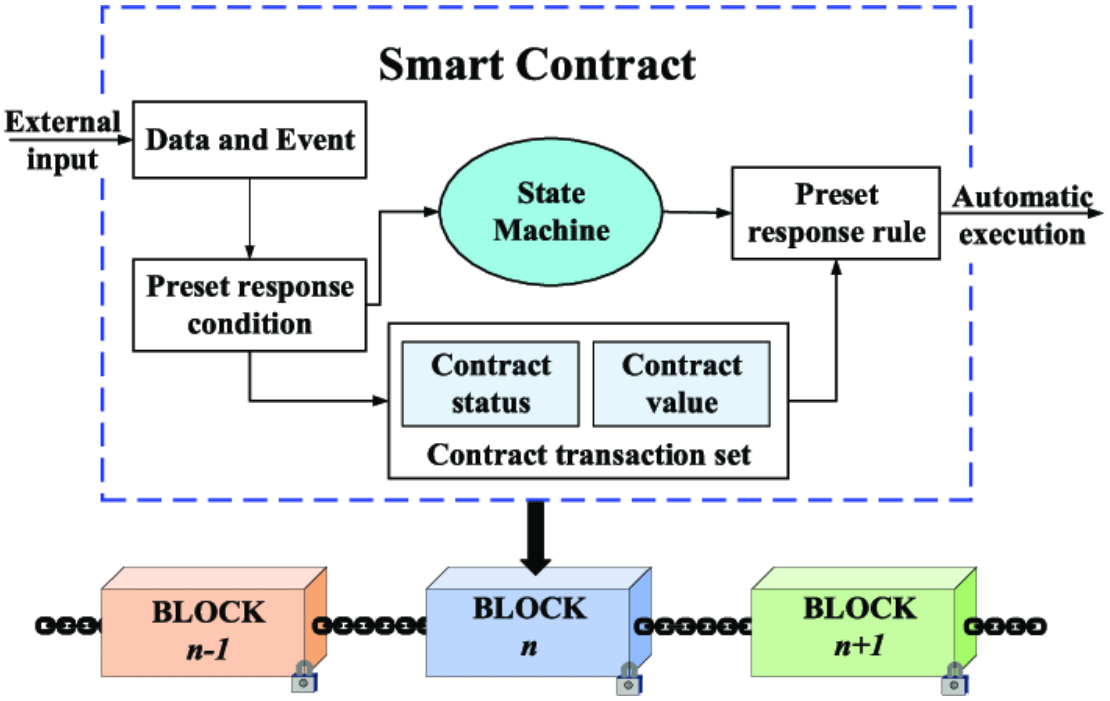

FIGURE 5. Smart contract model based on blockchain.

Smart contracts can be defined as a block of programming code containing streams of if-else statements Smart contracts are programmed according to the agreements between the contractors, signed cryptographically and then entered to the blockchain [18]. They have a view of the whole blockchain so they can automatically execute the contract terms defined in it. This contract is embedded in the blockchain

making it transparent, immutable, inexpensive and decentralized. Every smart contract has its address in the blockchain. The contract can be interacted with by using its address presuming the contract has been broadcasted in the network.

One important feature of using smart contracts is trust, since the smart contract cannot be lost as it is embedded in the blockchain itself. Great accuracy is also achieved, if the code that is destined for execution is precise. In addition, due to the fact that everything is coded, speed is a strong feature of smart contracts. They reduce the time it takes to maneuver through all the human actions that should have been done. The time it takes to do all the actions is equal to the time that the code in the contract takes to execute. There is also no need for backup, since every node in the blockchain maintains the shared ledger, providing one of the best backup facilities. There is also high safety, since the integrity of the contract is protected by the blockchain's cryptography. Even if someone breaks the encryption, the hacker will have to modify all the blocks that come after the block which has been modified. Please note that this is a highly difficult and computation intensive task and is practically impossible for a small or medium sized organization to do.

On the other hand, smart contracts come with some disadvantages. One of the most important is that bugs might exist in the code. An exploited bug can lead to breaking the terms of a contract, and the victims might now realize it immediately. The impossibility of changing something in the smart contract can also be a disadvantage. Fixing errors and changing contract terms will be an unsolvable problem. Since they are practically immutable, whenever there is a change that has to be incorporated into the contract, a new contract has to be made and implemented in the blockchain.

Overcoming CTI issues with Blockchain Technology

Cybersecurity incidents occur all the time inside an organization. Taking this fact to a global scope, there are a huge number of incidents, either significant or of less importance. When those incidents are documented in a CTI report, the number of those reports can grow dramatically. Without a proper, well thought way to organize those reports, a cyber threat intelligence consumer will find it hard to search for the right report that contains the intelligence that he is looking for. Blockchain can be very useful in managing and storing CTI reports. A good practice would be to store the CTI report data in a block. For each new CTI report, a new block will be added in the blockchain. In an opt-in scenario, the CTI producers will be able to report and distribute CTI reports between participants or subscribers to the blockchain. In addition, CTI consumers would be able to track changes and threats more easily.

Moreover, data integrity is crucial in cyber threat intelligence. Nowadays, there are multiple platforms that are used for CTI sharing. Many of them are based on open source intelligence, like the OpenCTI platform [21]. Users that have access to such platforms, can intentionally or unintentionally add, delete or modify the information stored in the database. This reduces the reliability of the CTI sharing platform. Reliability is one of the most important factors when making actionable intelligence. Organizations will only automate processes when they know that the information that they have in their hands is reliable. For example, if a SIEM tool is automatically executing queries based on cyber threat intelligence, the organization must not only have complete trust towards the producer of the CTI, but also towards the servers and methods that are used for storing. An attacker with access to the CTI sharing platform server's database, could modify IoCs to his or her advantage, by making the malicious activity pass undetected from the organization's systems that use the breached CTI platform's intelligence. This issue can be overcome with the use of a blockchain. Blockchain's strong cryptographic security which was discussed in the previous section, makes the modification of CTI data extremely difficult. A CTI report that is stored in a block, has its hash computed and stored inside the block and also in the previous block which works as a pointer, to create a link between

consecutive blocks. Because blockchain is a decentralized ledger, if someone tries to alter data, the system will analyze the entire chain, compare it to the data changes, and effectively kick out any changes that don't match up. As a result, blockchain implements the system's integrity security and prevents any unauthorized changes.

Proposed System

This section presents the architecture and core elements of the proposed BRIDGE CTI sharing platform (BRIDGing the gap bEtween CTI production and consumption). Our approach is focused on solving two major issues in CTI, the interoperability and the sharing issue, while maintaining the integrity of the threat intelligence.

Architecture

The overview of the BRIDGE system architecture is demonstrated in Figure 6.

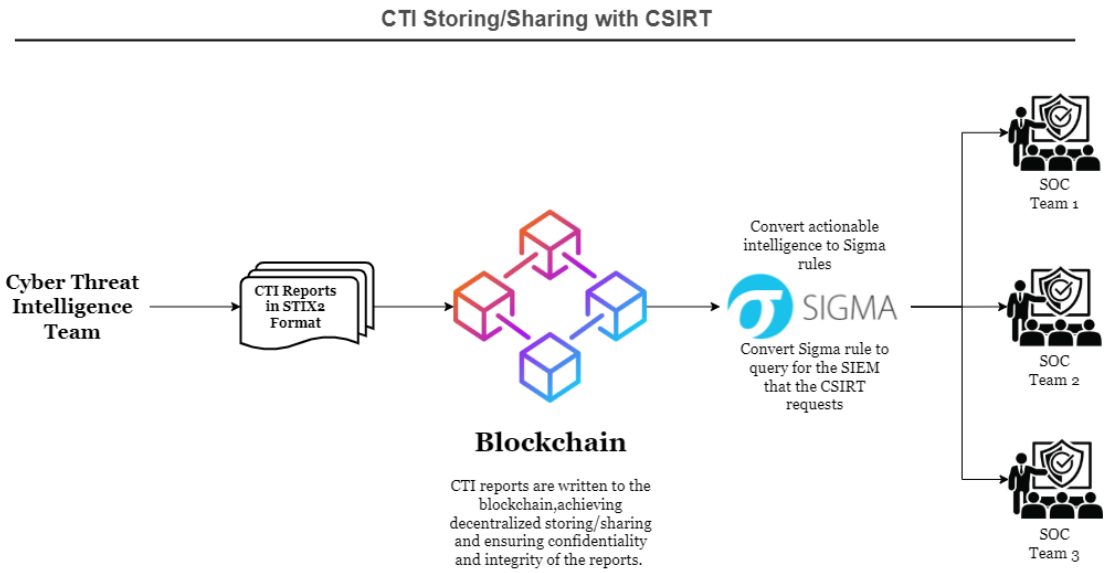

Figure 6. BRIDGE System Architecture

The input to this system is provided by the CTI team, which creates CTI reports that will be stored in the database. The data inside the block must follow a single CTI standard in order to allow the automations in the system's pipeline to execute successfully. Those automations are described later in this section. For that reason, the selected format of the report is the STIX 2.0 language standard.

In our proposed system, blockchain technology plays the role of the database. Both consumers and producers of threat intelligence are allowed to interact with the blockchain. Each time the CTI team wants to store a threat intelligence report, a new block is added to the blockchain containing the information inside the report.

On the other hand, the output of the system is requested by CTI consumers, which could be SOC teams that lead an investigation based on the actionable intelligence stored in the blockchain. When a CTI consumer requests to get a report, except from the report, a Sigma file is also supplied which describes the indicators found in the corresponding report. This Sigma file is a description of the detection method that an analyst should follow to detect the IOCs that are included in the CTI report.

The Sigma detection rule is vendor agnostic. With such a rule in hand, the SOC team can automatically generate a query to search for those indicators specifically crafted for the SIEM that they are using. Apart from the CTI report, the SIEM that the team uses for investigation can also be specified in the request. By supplying this, any actionable intelligence found in the form of indicators inside the report, will be returned inside a query for the desired SIEM.

Workflow

This section gives a high level overview of the operation of the proposed system. The flow can be described in two main phases, CTI production and CTI consumption. The steps for CTI sharing in both phases are described in order to present thoroughly the procedure followed by our architecture.

The first phase is the CTI production. The CTI team has gathered intelligence about threat actors, attacks and malware which want to share with other teams via our platform. That information should be saved in STIX 2 format. In order to follow a common standard in CTI sharing, we chose the STIX 2 language to be the format in which the threat intelligence will be stored and shared. After the CTI team constructs its STIX 2 reports, the reports are ready to be stored in the blockchain. A new block

is added to the blockchain for every new report that is being published, maintaining its integrity while being available for all the members in the blockchain.

The second phase is the CTI consumption. In this phase, various SOC teams can search for CTI reports inside the blockchain database. At this point it can be assumed that each team uses a different SIEM for event searching. After the CTI report of interest has been found in the database and the request to retrieve this report is being made, the CTI consumer is given the ability to select in which SIEM the threat intelligence will be searched upon. Moreover, if the SIEM option is supplied, two more items will accompany the STIX 2 report. The first one will be a Sigma rule that matches the indicators inside the report, and the second one will be a text file containing the query that searches for those indicators for the SIEM that was requested.

Summing up the process of CTI sharing in the platform proposed in this research, the flow consists of the following steps:

- The CTI team constructs reports in STIX 2 language
- Each new CTI report is stored in a new block in the blockchain
- SOC team members request for a CTI report and specify the SIEM that they use
- The CTI report, alongside with the Sigma rule that matches the intelligence inside the report and the query to search for the indicators on the specified SIEM is provided.

After the report has been received from the SOC team, not only the team members have saved time creating those queries manually, but also human errors that can lead to malformed queries are avoided. The analyst will be certain that the query will match all the indicators in the CTI report since it has been created based on the Sigma rule containing the threat intelligence. Moreover, multiple teams investigating the same incident that may work on different SIEM have overcome the interoperability issue. SOC teams will now be certain that they search for the same indicators, in the same way as the rest teams, since the queries that they use will be generated based on the same CTI report and Sigma rule.

Experimental Testing and Evaluation

In this section the testing environment is presented. Moreover, the testing scenarios will be described and how they were executed. While performing multiple scenarios of CTI sharing and utilizing the full potential of the pipeline, performance measurements will be taking place. Currently, most SOC teams cooperate with the CTI team in a more non-standardized, manual way. After conducting a survey on how a CTI sharing performs with traditional communication, the results will be compared against the performance measurements collected in the Testing section.

Lab Environment

All the experimentation is conducted in a Linux Ubuntu 20.04.4 [22] machine with x86_64 architecture. The system is using an Intel i5 10600K processor with 6 cores that support hyperthreading. The total available memory is 16GB and the disk storage is 500GB.

For the implementation, all code is written in Python programming language version 3.8.10 [23]. The CTI reports that were parsed by Python scripts strictly follow the STIX language standard, version 2.1 [17]. After parsing the STIX reports, the **sigmac** executable [24] is being run from Python scripts to generate the query for the desired SIEM.

Testing and Evaluation

In order to test the performance of this implementation, multiple CTI reports in STIX 2.1 format have been created. For the first testing scenario, the number of IOCs in the reports is increased every time the code that creates the SIEM queries is being run, while the number of SIEM queries requested from each report remains the same. The goal of those tests is to measure the time consumed solely for parsing the indicators. The results are shown in Fig 7.

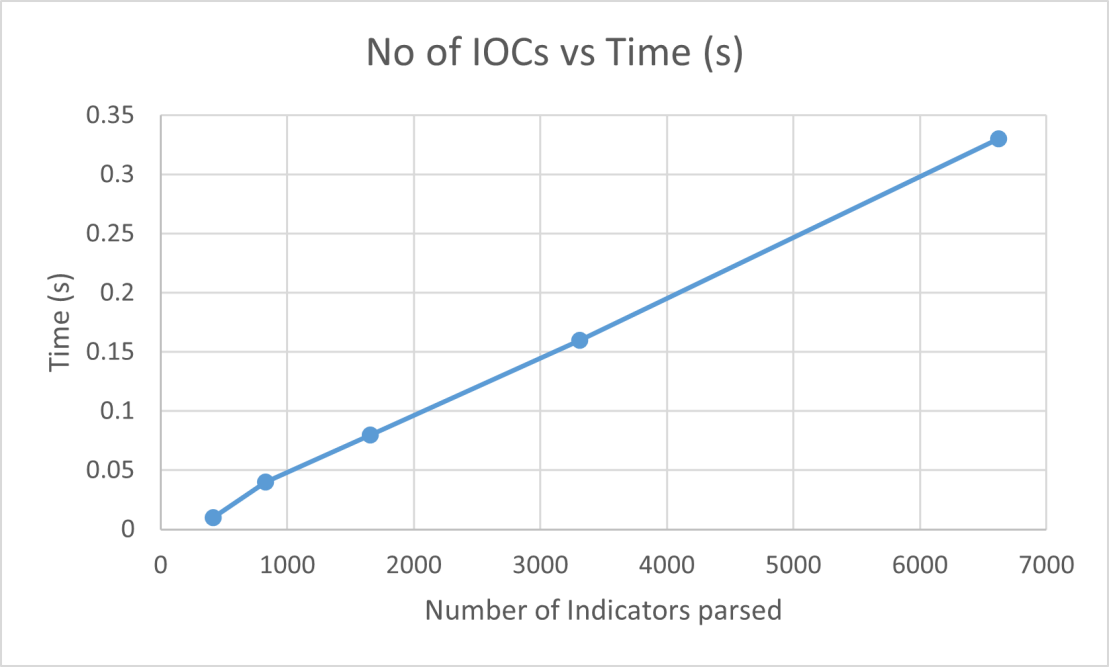

Fig 7. Number of indicators parsed over time

Reviewing the previous plot, the time taken to parse indicators demonstrates a linear increase. By doubling the amount of indicators inside the CTI reports, the time it takes to create the queries is also doubled. However, the time needed to parse the indicators is very minimal, taking only one second to parse around 20.000 indicators.

The next series of tests focus on measuring the time consumed for creating multiple SIEM queries from a CTI report. The number of indicators remains the same in all tests, while the number of SIEM queries is increased. The results are presented in Figure 8.

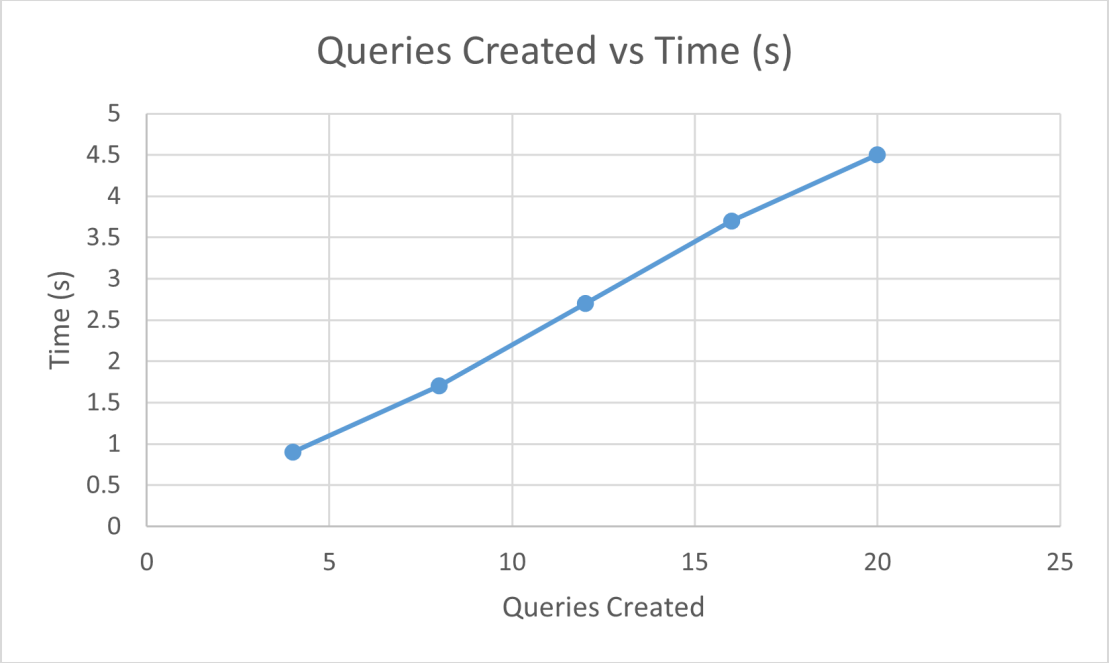

Fig 8. Number of queries created over time

The previous plot shows that the time consumed to create the queries is increasing linearly. It is observable that by doubling the number of queries requested, the code takes double the time to finish execution. By comparing Fig X with Fig X, it is concluded that the time consumed for parsing the indicators is negligible while compared to the time taken to generate the SIEM queries. This means that the CTI sharing platform will take more load by requesting queries for multiple SIEMs rather than requesting multiple indicators on a single SIEM query.

To quantify more accurately the impact of BRIDGE in a real environment, similar experiments have been conducted in a working environment consisting of 10 security analysts. The time each analyst would need to create a query for the SIEM that they use was measured. Creating SIEM queries based on a CTI report is an everyday task which is very important for monitoring the security of the infrastructure that they are responsible for. The following table sums up the time consumption of generating the SIEM query based on CTI reports with our proposed BRIDGE architecture, while on the other hand it presents the time it would take for the query to be created manually. The time was tracked separately for each analyst the the results

are the mean time consumed among 10 security analysts participated in this survey. They executed one query to the CTI report (used before) fetching 5, 10, 15 and 20 IOCs (see Table I).

Evaluated Method	**# of IOCs fetched per query**	**Time (sec.)**
BRIDGE	414 828 1656 3312 6624 13248 19872	0.01 0.01 0.08 0.16 0.33 0.68 0.91
Traditional SOC method	5 10 15 20	70.2 182.4 247.2 274.2

TABLE I. Fetching numerous IOCs of one CTI report

The comparison proved that the traditional way that SOC teams process their daily routine has become rigid, while the cyber security needs are on rise; however, our implementation is able to fight this rigid way providing effectiveness and speed maintaining the quality that is required in these critical tasks.

Conclusion

In this research we first investigated how the threat intelligence has begun to be utilized and how it evolved to be playing a critical role in security monitoring and incident response. We studied the CTI lifecycle and what are the steps that should be performed in each phase of it. Moreover, we examined the role of CTI in Security Operations Centers, how they operate together and what are the issues and challenges that arise.

Following the detection of the problems that security professionals face, we present the key features of blockchain technology and how they can be helpful in overcoming those challenges. Finally, we propose the implementation of BRIDGE, a CTI sharing platform with the ability to automatically generate SIEM queries according to the IOCs included in the CTI report that needs to be investigated. After the detailed analysis of the system's architecture, the performance of this platform is tested and evaluated. It is also compared to the performance measurements of current traditional SOC methods for CTI sharing and for SIEM queries creation, which proves that the BRIDGE CTI sharing platform aids to advance the CTI lifecycle many iterations ahead.

Contribution

In this study the first CTI sharing tool is presented, which is focused on the automation of the information consumption phase, specifically designed for cybersecurity professionals and practitioners. Evaluating BRIDGE, it is proved that the beneficiaries and especially SOC teams can take advantage of BRIDGE to automatically create queries for their SIEM and at the same time eliminate human errors, enable interoperability via the STIX format and Sigma rules, and establish a transparent method for managing security incidents. The aforementioned benefits are only the technical advantages that follow BRIDGE; however, the integration of BRIDGE to the arsenal of CTI consumers can also increase the quality of security decisions taken from all CTI consumers. At the core of the BRIDGE tool lies the integration of the STIX standard, which offers indisputable interoperability and

creates a common expression within the CTI ecosystem. Having designed and developed the BRIDGE tool, its performance is quantitatively evaluated and proved that it is able to successfully cope with the current issues that SOC members meet in their working routine. As the number of security incidents and challenges are on the rise, more security information will be produced by the CTI mechanism and new SIEM tools will emerge. It is concluded that the BRIDGE research outcomes will pave the way for a CTI ecosystem armed with a unified expression to fight back and defend against various critical cybersecurity threats. It is also expected that BRIDGE will be the precursor for an automated CTI ecosystem being able to address the numerous cybersecurity threats that daily emerge. Additionally, more Threat Intelligence Sharing Platforms start producing CTI reports in STIX format and together with the integration of the BRIDGE tool can achieve automation and high success-levels in security incidents handling.

Future Work

The research outcomes of this thesis can be extended as future work in many ways. For this proof-of-concept implementation of BRIDGE, a prototype for Unix-based environments was designed and developed. Future plans include implementing BRIDGE for Windows based environments removing environment-related barriers. In addition, future work includes developing and integrating a Self-Sovereign-Identity approach within blockchain technology to create an ecosystem with trustworthy CTI consumers, who may belong to different organizations but should share their intelligence and security information increasing. Also, the list of SIEM that Sigma supports could be enhanced to increase interoperability.

References

[1]Watkins K-F. M-Trends 2017: A view from the front lines. Vol. 4, Premier Outlook. 2017

[2]Kaur Sahi Asst S. A Study of WannaCry Ransomware Attack. Int J Eng Res Comput Sci Eng. 2017;4(9):7–9.

[3]“Central Intelligence Agency,” Visit the main page. [Online]. Available: https://www.newworldencyclopedia.org/entry/Central_Intelligence_Agency. [Accessed: 15-Apr-2022].

[4]D. Shiloach, Doron Shiloach X-Force Product Manager, D. Shiloach, X.-F. P. Manager, and Doron Shiloach is a product manager with the IBM X-Force threat and vulnerability research team, “The path forward with threat intelligence and sharing,” Security Intelligence, 13-Jan-2016. [Online]. Available: https://securityintelligence.com/threat-intelligence-and-sharing/. [Accessed: 15-Apr-2022].

[5]M. Waite, Pocket oxford english dictionary. Oxford: Oxford University Press, 2013.

[6]“What is Cyber Threat Intelligence? [beginner's guide],” crowdstrike.com, 17-Mar-2022. [Online]. Available: https://www.crowdstrike.com/cybersecurity-101/threat-intelligence/. [Accessed: 15-Apr-2022].

[7]“How to empower your SOC with security intelligence,” Recorded Future, 21-Jan-2020. [Online]. Available: https://www.recordedfuture.com/empower-security-operations/. [Accessed: 15-Apr-2022].

[8]J. Oltsik, “Dealing with overwhelming volumes of security alerts,” ESG, a division of TechTarget. [Online]. Available: https://www.esg-global.com/blog/dealing-with-overwhelming-volume-of-security-alerts. [Accessed: 15-Apr-2022].

[9]“Siem Productivity Report,” Exabeam. [Online]. Available: https://www.exabeam.com/library/exabeam-siem-productivity-report/. [Accessed: 15-Apr-2022].

[10]Cisco, "Annual Cybersecurity Report", 2018.

[11]T. Schaberreiter, V. Kupfersberger, K. Rantos, A. Spyros, A. Papanikolaou, C. Ilioudis, and G. Quirchmayr, “A quantitative evaluation of trust in the quality of Cyber Threat Intelligence Sources,” Proceedings of the 14th International Conference on Availability, Reliability and Security, 2019.

[12]P. Institute LLC, The Value of Threat Intelligence: Annual Study of North American & United Kingdom Companies, Jul. 2016.

[13]Tara Seals US/North America News Reporter, “Threat intelligence strategies suffer from data overload,” Infosecurity Magazine, 14-Sep-2017. [Online]. Available: https://www.infosecurity-magazine.com/news/threat-intelligence-strategies/. [Accessed: 15-Apr-2022].

[14]X. Liao, K. Yuan, X. F. Wang, Z. Li, L. Xing, and R. Beyah, “Acing the IOC game,” Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016.

[15]D. Homan, I. Shiel, and C. Thorpe, “A new network model for cyber threat intelligence sharing using blockchain technology,” 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 2019.

[16]“Cyber observable expression (cybox™) archive website,” CybOX. [Online]. Available: https://cyboxproject.github.io/. [Accessed: 15-Apr-2022].

[17]“Structured threat information expression (STIX™) 1.x archive website,” STIX. [Online]. Available: https://stixproject.github.io/. [Accessed: 15-Apr-2022].

[18]E. BÜBER and Ö. K. ŞAHİNGÖZ, “Blockchain based information sharing mechanism for Cyber Threat Intelligence,” Balkan Journal of Electrical and Computer Engineering, vol. 8, no. 3, pp. 242–253, 2020.

[19]D. Homan, I. Shiel, and C. Thorpe, “A new network model for cyber threat intelligence sharing using blockchain technology,” 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS), 2019.

[20]J. Yli-Huumo, D. Ko, S. Choi, S. Park, and K. Smolander, “Where is current research on blockchain technology?—a systematic review,” PLOS ONE, vol. 11, no. 10, 2016.

[21]“Open platform for cyber threat intelligence,” OpenCTI, 13-Mar-2022. [Online]. Available: https://www.opencti.io/en/. [Accessed: 16-Apr-2022].

[22]“Download ubuntu desktop: Download,” Ubuntu. [Online]. Available: https://ubuntu.com/download/desktop. [Accessed: 16-Apr-2022].

[23]“Welcome to Python.org,” Python.org. [Online]. Available: http://www.python.org/. [Accessed: 16-Apr-2022].

[24]“Sigma,” GitHub. [Online]. Available: https://github.com/SigmaHQ. [Accessed: 16-Apr-2022].