BRIDGE: BRIDGing the gap bEtween CTI production and consumption

Marios Karatisoglou
Department of Digital Systems
University of Piraeus
Piraeus, Greece
m.karatisoglou@ssl-unipi.gr

Aristeidis Farao
Department of Digital Systems
University of Piraeus
Piraeus, Greece
arisfarao@unipi.gr

Vaios Bolgouras
Department of Digital Systems
University of Piraeus
Piraeus, Greece
vbolgouras@unipi.gr

Christos Xenakis
Department of Digital Systems
University of Piraeus
Piraeus, Greece
xenakis@unipi.gr

2022 14th International Conference on Communications (COMM) | 978-1-6654-9485-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/COMM54429.2022.9817359

***Abstract*—Security for businesses and organizations is essential to protect operational activities, trust relationship with 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 infrastructures. Automation of procedures regarding the collection, harmonization and processing of information is of utmost importance for Cyber Threat Intelligence, in order to effectively relay to the community data concerning newly emerged threats. Nevertheless, the process regarding the transfer of knowledge between Cyber Threat Intelligence 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 Cyber Threat Intelligence and cybersecurity professionals, by taking advantage of the Structured Threat Information eXpression 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.**

***Index Terms*—CTI, Automation, Sharing, Interoperability**

I. Introduction

Threat intelligence is rapidly becoming a priority for businesses and organizations across the globe, due to the continuously emerging modern cyberattacks and their sophistication level [1]. Malicious actors performing criminal activities in the cyberspace showcase exceptional skills in their tactics, techniques and procedures, thus it becomes exceedingly difficult and challenging for cybersecurity professionals to investigate and intercept their activity [2]. Cybersecurity researchers and professionals working in environments like the Security Operations Centers (SOCs) [3] are employed in order to find ways to mitigate threats and monitor large amounts of data pertaining to organizations' infrastructures. To that end, cybersecurity tools like firewalls, intrusion detection and prevention systems and Security Information and Event Management systems (SIEMs) are utilized.

To battle the never-ending stream of newly emerged cyberattacks and swiftly update the network among the cybersecurity professionals, Cyber Threat Intelligence (CTI) programs are being widely employed [4]. Through CTI, the community can be up to date regarding existing threats and attacks that have already taken place at least once, giving them the ability to proactively mitigate advanced threats. CTI is a fundamental concept that exists since the early days of cybersecurity's adoption by numerous organizations and institutes, which has evolved according to the advancements that have occurred in this sector. The large scale security event data that is created, the need for swift analysis and processing of intelligence and the never ending growth of the threat landscape, has resulted in the automation of almost all the CTI procedures - intelligence gathering, processing of information and harmonization of reporting.

While CTI infrastructures are vital for researchers and cybersecurity professionals to perform their duties efficiently and minimize risks, there are a few shortcomings [5]:

- CTI consumers (researchers, cybersecurity professionals e.t.c.) have access to intelligence that has been gathered, processed and reported by automated means, but still have to **manually** extract the information needed in order to use it for research or threat mitigation purposes.
- There has been a significant growth in the number of threat data sources, from which a CTI practitioner has to generate useful intelligence that can be used in decision-making processes. 70% of respondents in [6] declared that threat intelligence is too **voluminous and/or complex** to provide actionable intelligence. Unfortunately, companies are collecting massive amounts of data in a wide variety of different formats such as Structured

978-1-6654-9485-4/22/$31.00 ©2022 IEEE

Threat Information eXpression standard (STIX), JSON, XML,PDF, CSV, email without keeping a standard format hardening CTI consumers to manually processes and review the gathered data.

- Lastly, CTI consumers [3] in their effort to mitigate threats [7], [8] must control all the data created by the growing number of data locations and sources. This undertaking becomes increasingly complex because of the variety of security measures and tools utilized for this purpose. As a result, it is vital to establish standards and procedures that ensure **interoperability** among these components, facilitating security operations and response procedures throughout the whole security ecosystem.

We solve the aforementioned challenges with BRIDGE, a novel implementation and to the best of our knowledge the first effort to bridge the gap between CTI and its consumers, by automating the process of converting information stemming from CTI reports to the format needed by the researchers and cyber security professionals. Employing the STIX standard to store information at CTI reports, BRIDGE gives the ability to the CTI consumers to automatically apply the information provided on a variety of tools. No further manual input or modification is required. Moreover, the blockchain technology is also utilized to safely store in a single but decentralized repository all the CTI data, which gives the ability to the professionals to easily monitor the information that is provided and ensure certain level of quality, as they are not required to oversee numerous repositories.

The overall contribution of this work is the following:

- implement a decentralized CTI sharing platform based on blockchain
- CTI consumers can automatically generate data in the desired format for their tools, based on indicators provided by CTI reports
- CTI lifecycle has progressed significantly by filling the gap between CTI and its consumers.

The paper unfolds as follows: Section II presents essential background information on CTI ecosystem and information regarding related works that propose solutions towards the automation of CTI procedures. Next, Section III elaborates on the processes of the BRIDGE tool describing in detail all the required steps. Section IV includes a quantitative performance evaluation of BRIDGE, and Section V concludes the paper.

II. Background

This Section presents the CTI [9] concept and describes how cybersecurity researchers and professionals utilize the CTI infrastructure, improving the defense against threats.

A. The CTI concept

The implementation of CTI follows a defined lifecycle that consists of seven discrete phases: (i) requirements; (ii) collection, processing; (iii) analysis; (iv) dissemination, (v) consumption and (vi) feedback. This flow ensures that CTI actions are in line with the organization's goals and produce actionable data with the appropriate meanings. Following this lifecycle, an organization can achieve constant improvement, which is one of the most important aspects in order to keep the CTI productive and effective.

Requirements: Threat intelligence's initial phase is responsible to establish the goal and scope of all intelligence actions. Also, it identifies the information assets and business processes that need to be protected, alongside with the potential impacts of losing those assets or interrupting those processes. These have been prioritized according to what is more important to protect. Finally, this phase defines the possible attackers, their actions and their motivation.

Collection: Once the requirements are defined, the CTI team will seek to collect the required data to achieve those objectives. On the one hand, internal sources will be exploited such as metadata and traffic logs from internal networks and devices. On the other hand, external sources will be utilized such as scrapping and crawling dark web forums and open source intelligence databases, as well as human intelligence will be investigated [10].

Processing: Processing entails converting raw data, that came from the *Collection* phase, into a format suitable for further investigation and analysis, e.g., harmonization. Processors might be either humans or robots executing specific algorithms depending on how the data was collected.

Analysis: After the raw data is processed in the aforementioned step, the CTI team will undertake a comprehensive analysis to meet the goals set in the initial phase, also its outcome is a report summarizing the security data. In particular, artificial intelligence, data analytics and machine learning are utilized by the CTI team to make predictions and extract insights and patterns, to analyze raw data to make conclusions, as well as to predict and find representative values for the missing data

Dissemination: Dissemination entails delivering the completed intelligence product to the appropriate audience. First and foremost, this phase identifies the detected threats. Once the identification is completed the organization's cybersecurity status is evaluated and the most optimal strategies and security controls are proposed to strengthen the organization to defend future cybersecurity threats. The proposed security solutions include but are not limited to risk transfer, installation of security tools as well as compliance with standards [11].

Consumption: CTI consumers receive the data from the corresponding repositories, which then has to be processed in order to meet the requirements of the tools and technologies that will utilize it. This step can be time consuming and because of the lack of automated means, the manual process that is carried out may result in the corruption of information.

Feedback: This is the last phase, that is responsible to assess on a continuous basis the cybersecurity level of the organization as well as the performance of the implemented cybersecurity controls.

We can observe that CTI is not a process with start and end point, but it is a loop consisting of phases that feed off each other.

B. Related work

Numerous works related to the CTI concept are focused on the enhancement and improvement of the performance concerning the corresponding procedures followed by CTI producers. Like BRIDGE, automation of processes is the key for the majority of solutions which focus on many of the aforementioned phases. Below we mention indicatively some works that aim to automate CTI procedures.

For the collection of data, the authors of [12] propose an automatic approach to generate the CTI records, which is based on the Natural Language Processing (NLP) and machine learning concepts. Other efforts focusing on the automatic collection and processing of can be found in the works of [13] and [14], where Indicators Of Compromise (IOCs) are extracted from the corresponding data. To both process and analyse the collected data, the authors of [15] propose a solution that processes Malware Information Sharing Platform (MISP) data automatically, prioritizes cybersecurity threats for Small and medium-sized enterprises (SMEs), and provides SMEs with actionable recommendations tailored to their context. On top of solutions like the ones mentioned above, in order to facilitate the automation of processes regarding the CTI, standardization efforts have been made. STIX [16], which is also utilized by BRIDGE, is considered the main standard that should be adopted in order to describe threat intelligence data and be used by threat intelligence sharing platforms [17].

All solutions regarding the automation of the procedures found in CTI's lifecycle focus on the phases of collection, processing and analysis of corresponding data. Regarding the Consumption phase of the produced intelligence there had been no efforts so far. This gap will be filled with BRIDGE, a solution that aims in automating the process of converting data that stems from CTI infrastructures to the form that each consumer needs it.

III. BRIDGE

A. Software architecture

BRIDGE aims to greatly facilitate from top to bottom members of CTI consumption ecosystem, including but not limited to SOC teams, Security and Information Technology Analysts (Sec/IT Analysts), Computer Security Incident Response Teams (CSIRT), Intelligence Analysts, Board of Directors and Security Researchers. Through BRIDGE, sharing CTI results among the aforementioned parties will result in establishing interoperability, maintaining the integrity of the produced CTI information and create the ideal conditions to effectively extract crucial intelligence.

As shown in Fig. 1 the general structure of the BRIDGE is divided into three main modules: i) the Parser; ii) the Translator, and iii) the Data Pool. We have to note that for demonstrative purpose and ease of understanding, BRIDGE is presented assuming that SOC teams will be the end consumers - however, we avoid analyzing SOC processes since it is out of scope of this work. The **Parser** module as its name implies, is responsible to receive the CTI reports in STIX 2.1 [18] format standard generated by the corresponding team. The STIX 2.1 report is stored to the **Data Pool**. The latter is built based on blockchain technology and plays the role of the database providing the information system with immutability, integrity, transparency, and traceability of data shared across the organization network. We have to note, that each time the CTI team aims to store a report, a new block is added to the blockchain containing the information of the corresponding report. Then the **Translator** is getting requests from the organization SOC teams that manage different SIEM tools. Each SOC team requests from the **Translator** to get a CTI report, then a Sigma file is supplied describing the indicators found in the corresponding report. By utilizing Sigma files, which include Sigma rules, SOC teams are able to describe relevant log events in a flexible and standardized format. More specifically, the aforementioned report contains a description of the detection method that a SOC member 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 arsenal, 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.

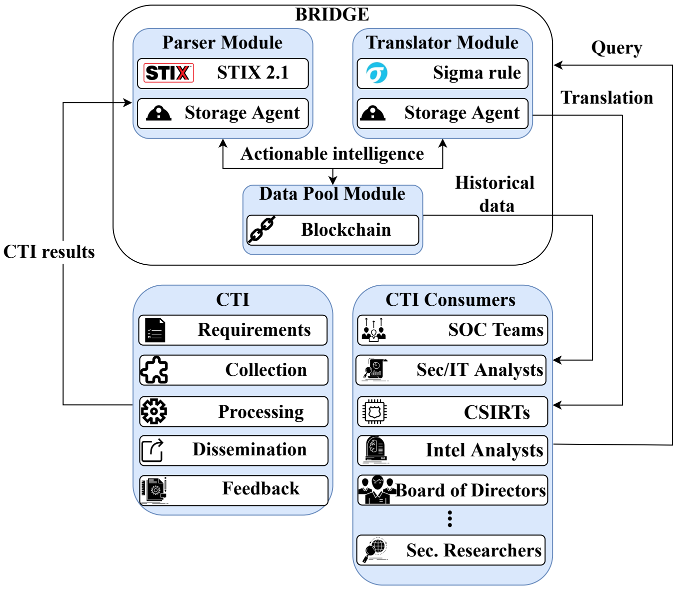

Fig. 1. BRIDGE architectural components

B. Technical approach and methodology

In this section we analyze how BRIDGE operates providing a workflow that should be followed. In particular, BRIDGE consists of two discrete phases: i) CTI production and ii) CTI consumption.

The first phase entitled **CTI production** is responsible to receive the generated CTI report. First and foremost, the CTI team gathers intelligence about threat actors, cyberattacks and malware, which will be shared with SOC teams, and store them under STIX 2.1 format via the BRIDGE parser (see Fig. 1). 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 legitimate members in the blockchain.

After the successful completion of the first phase (CTI production), **CTI consumption** is being performed by the SOC teams. The latter utilize various SIEM tools for event investigation. After the CTI report of interest has been found within the blockchain 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. The key element of a CTI report is the IOCs that constitute it. Moreover, once the SIEM option is supplied, two more items will accompany the STIX report. The first one will be a Sigma rule that matches the IOCs inside the report, and the second one will be a text file containing the query that searches for those indicators for the particular SIEM.

Overall, after the CTI report has been delivered at the SOC teams, the latter not only have saved time creating SIEM queries automatically via BRIDGE, but also human errors that can lead to malformed queries have been eliminated due to the acceleration that BRIDGE provides incorporating that Sigma. In addition, the involved analysts utilizing the BRIDGE tool are certain that the query matches all the IOCs in the CTI report since it has been created based on the Sigma rule containing the threat intelligence. Finally, multiple teams investigating the same incident that may work on different SIEM have overcome the interoperability issue.

IV. Performance Evaluation

In this Section, we analyze the performance of the tool as a whole, investigating its feasibility and efficiency against incidents that come from real working environment. BRIDGE has been developed in Python (version 3.8.10) language, utilizes the CTI reports strictly following the STIX 2.1 language standard, version 2.1, finally the dedicated SIEM rule is generated by the sigma rule. The experiments have been conducted in a Ubuntu Desktop 20.04.4 being equipped with an Intel i5-10600K processor with 6 cores that support hyperthreading at 4.1 GHz, 16GB RAM and 500GB disk storage. By now, BRIDGE runs in Unix-based operation systems (i.e. Linux). We have conducted two experiments to evaluate BRIDGE efficiency and to compare its effectiveness against the method that is currently being used by the SOC teams, which has been chosen to represent the consumer of BRIDGE in our evaluation. SOC is among the top professional groups that will utilize the produced CTI reports (see Fig. 1), while at the same time they lack proper and automated bridge hub solution in order to fetch IOCs for their SIEMs. We have to note that the conducted experiments aim to evaluate the performance of BRIDGE's main modules; however, the blockchain component is utilized as a database and does not have any impact on the performance of the BRIDGE tool, which is evaluated after the data has already been fetched. Thus, no measurements regarding the blockchain infrastructure's performance were taken.

The first experiment aims to measure the time consumed solely for parsing the IOCs from a CTI report. In particular, we generate one query for the Splunk SIEM [19] that is executed many times against one CTI report. Each query fetched 414, 828, 1656, 3312, 6624, 13248 and 19872 IOCs. The experiment was conducted 5 times. The produced results revealed that our tool required less than a minute to fetch thousands of IOCs from one CTI report (see Table I). The same experiment has been conducted by the assistance of 10 professional cybersecurity analysts, who are members of SOC teams (they voluntarily participated). They executed one query to the CTI report (used before) fetching 5, 10, 15 and 20 IOCs (see Table I). The comparison proved that the traditional way that SOC teams process their daily routine has became rigid, while the cybersecutiy needs are in 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.

TABLE I
Fetching numerous IOCs of one CTI report

Evaluated method	**# of IOCs fetched per query**	**Time (sec.)**
BRIDGE	414	0.01
	828	0.04
	1656	0.08
	3312	0.16
	6624	0.33
	13248	0.68
	19872	0.91
Traditional SOC method	5	70.2
	10	182.4
	15	247.2
	20	274.2

For the second experiment completion, we used one CTI report with many IOCs and executed numerous queries at the same time for 4, 8, 12, 16, 20, 24 and 28 different SIEM tools. Each query fetches the same 44 IOCs (see Fig. 2). Also, the experiment was conducted 5 times. We can observe that time needed to create queries for different SIEM fetching standard number of IOCs increases linearly. Overall, we can validate that the time consumed for parsing the numerous IOCs for a specific SIEM is negligible compared to the time needed to generate the SIEM queries manually. Also, we can observe that BRIDGE performs better when requesting multiple indicators on a single SIEM query rather than requesting queries for multiple SIEMs.

V. Conclusions

In this paper we presented the first CTI sharing tool, which is focused on the automation of the information consumption phase, specifically designed for cybersecurity professionals and practitioners. Evaluating BRIDGE, we have proven that the beneficiaries and especially *SOC* teams can take advantage of BRIDGE to automatically create queries for their SIEM and

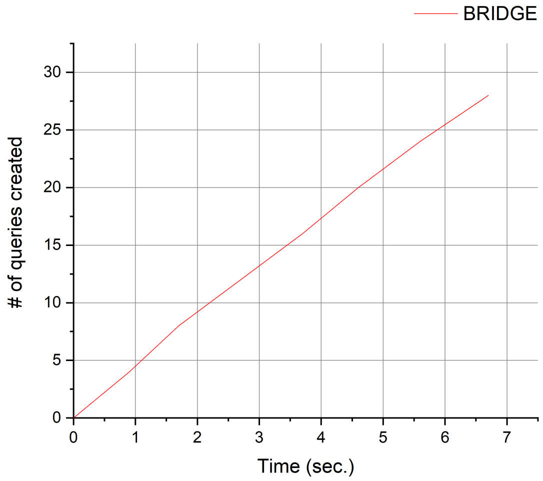

Fig. 2. Fetching numerous queries for numerous SIEM tools

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 them.

At the core of the BRIDGE tool we find the integration of STIX standard, which offers indisputable interoperability and creating a common expression within the CTI ecosystem. Having designed and developed the BRIDGE tool, we quantitatively evaluated its performance and proven 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 in the rise, more security information will be produced by the CTI mechanism and new SIEM tools will emerge. Our belief is 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. We also expect 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 BRIDGE tool can achieve automation and high-success-levels in security incidents handling.

The research outcomes of this paper can be extended as future work in many ways. For this proof-of-concept implementation of BRIDGE, we designed and developed a prototype for Unix-based environments. Next, we plan to implement BRIDGE for Windows based environments removing environment-related barriers. In addition, we aim to develop and integrate 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, we aim to enhance the list of SIEM that Sigma suports by increasing interoperability.

Acknowledgment

This research was funded by the European Commission, under the SealedGRID, SECONDO and INCOGNITO projects; Grant Agreements no 777996, 823997 and 824015, respectively.

References

[1] Aristeidis Farao, Sakshyam Panda, Sofia Anna Menesidou, Entso Veliou, Nikolaos Episkopos, George Kalatzantonakis, Farnaz Mohammadi, Nikolaos Georgopoulos, Michael Sirivianos, Nikos Salamanos, et al. Secondo: A platform for cybersecurity investments and cyber insurance decisions. In *International Conference on Trust and Privacy in Digital Business*, pages 65–74. Springer, 2020.

[2] Vaios Bolgouras, Christoforos Ntantogian, Emmanouil Panaousis, and Christos Xenakis. Distributed key management in microgrids. *IEEE Transactions on Industrial Informatics*, 16(3):2125–2133, 2019.

[3] Deepesh Shahjee and Nilesh Ware. Integrated network and security operation center: A systematic analysis. *IEEE Access*, 10:27881–27898, 2022.

[4] Xander Bouwman, Victor Le Pochat, Pawel Foremski, Tom Van Goethem, Carlos H Gañán, Giovane Moura, Samaneh Tajalizadehkhoob, Wouter Joosen, and Michel van Eeten. Helping hands: Measuring the impact of a large threat intelligence sharing community. In *Proceedings of the 31st USENIX Security Symposium*. USENIX Association, 2022.

[5] Md Sahrom Abu, Siti Rahayu Selamat, Aswami Ariffin, and Robiah Yusof. Cyber threat intelligence–issue and challenges. *Indonesian Journal of Electrical Engineering and Computer Science*, 10(1):371–379, 2018.

[6] Ponemon Institute LLC. The value of threat intelligence : A study of north american & united kingdom companies sponsored by anomali., 2016.

[7] Enoch Agyepong, Yulia Cherdantseva, Philipp Reinecke, and Pete Burnap. Challenges and performance metrics for security operations center analysts: a systematic review. *Journal of Cyber Security Technology*, 4(3):125–152, 2020.

[8] Manfred Vielberth, Fabian Böhm, Ines Fichtinger, and Günther Pernul. Security operations center: A systematic study and open challenges. *IEEE Access*, 8:227756–227779, 2020.

[9] Daniel Schlette, Marco Caselli, and Günther Pernul. A comparative study on cyber threat intelligence: The security incident response perspective. *IEEE Communications Surveys Tutorials*, 23(4):2525–2556, 2021.

[10] Christoforos Ntantogian, Panagiotis Bountakas, Dimitris Antonaropoulos, Constantinos Patsakis, and Christos Xenakis. Nodexp: Node. js server-side javascript injection vulnerability detection and exploitation. *Journal of Information Security and Applications*, 58:102752, 2021.

[11] Sakshyam Panda, Aristeidis Farao, Emmanouil Panaousis, and Christos Xenakis. *Cyber-Insurance: Past, Present and Future*, pages 1–4. Springer Berlin Heidelberg, Berlin, Heidelberg, 2019.

[12] Tianfang Sun, Pin Yang, Mengming Li, and Shan Liao. An automatic generation approach of the cyber threat intelligence records based on multi-source information fusion. *Future Internet*, 13(2):40, 2021.

[13] Shengping Zhou, Zi Long, Lianzhi Tan, and Hao Guo. Automatic identification of indicators of compromise using neural-based sequence labelling. *arXiv preprint arXiv:1810.10156*, 2018.

[14] Zi Long, Lianzhi Tan, Shengping Zhou, Chaoyang He, and Xin Liu. Collecting indicators of compromise from unstructured text of cybersecurity articles using neural-based sequence labelling. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019.

[15] Max van Haastrecht, Guy Golpur, Gilad Tzismadia, Rolan Kab, Cristian Priboi, Dumitru David, Adrian Răcătăian, Louis Baumgartner, Samuel Fricker, Jose Francisco Ruiz, et al. A shared cyber threat intelligence solution for smes. *Electronics*, 10(23):2913, 2021.

[16] Sean Barnum. Standardizing cyber threat intelligence information with the structured threat information expression (stix). *Mitre Corporation*, 11:1–22, 2012.

[17] Christian Sillaber, Clemens Sauerwein, Andrea Mussmann, and Ruth Breu. Data quality challenges and future research directions in threat intelligence sharing practice. In *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security*, pages 65–70, 2016.

[18] OASIS Cyber Threat Intelligence (CTI) TC. Stix™ version 2.1. https://docs.oasis-open.org/cti/stix/v2.1/csprd01/stix-v2.1-csprd01.html, Online: (last accessed on 06/04/2022).

[19] Splunk SIEM. https://www.splunk.com/, Online: (last accessed on 06/04/2022).