# Attack Data Usability and Challenges in its Capturing and Sharing

Ashima Rattan
CEC Landran,
Punjab

Navroop Kaur
P.E. CSTD
Cdac,Mohali

Saurabh Chamotra
Principal Engg.
Cdac,Mohali

Dr. Shashi Bhushan
HOD of IT
CEC Landran Punjab

**Abstract-** Now a days Attacker's launch attack campaigns targeting the zero day vulnerability, compromising internet users on a large scale. The first response to such campaigns is to detect them and collect sufficient information regarding tools, techniques used to exploit the vulnerability. Hence effective capturing of the attack data and its timely dissemination to defenders is required for the mitigation and prevention of the large scale attacks. The objective of our survey is to make an automated attack data capturing and sharing mechanism. This study will help us in finalizing the standard format for information sharing, further which could be machine readable. Such a system would be able to timely capture and identify the presence of a large scale targeted attack campaigns and effectively share the attack data with the security agencies in a format that is readily usable\actionable by them. We have also highlighted the fact that the format for sharing attack data is very crucial and the data sharing format should be machine digestible to reduce the human intervention and increase the response time.

**Keywords- Attack data; Cyber-attacks; STIX Extensions.**

## I. INTRODUCTION

Cyber-attacks are ever changing and are diverse in nature. They vary in their techniques, targets and motives. It is this diversity of the cyber-attack which makes them difficult to detect. Various detection tools (Antivirus, Intrusion Detection System, Intrusion Prevention System [1], Snort, Honeypots, Honeynets) [2], [3] and techniques are used to identify, stoppage and mitigation of the cyber-attacks. We can capture attack logs from these tools [4] apart from detecting and preventing cyber attacks. These attack logs contain information about the attacks prevented/detected/observed along with the timestamp. Many time logs of attacks are captured by the system which could not be fully detected by them but based on some features it is labeled as malicious. Such data collected over a period of time from various sources can be of great usage for organizations involved in threat intelligence gathering, prediction, and even useful in the detection of large scale infections. Online communities/organizations (i.e Dshield, CriticalStack, TeamCymru, AbuseHelper, autoshun) collects such threat intelligence, process it in form which is useful to the network security community [8]. Further there are tools both open-source and commercial (i.e Abusehelper, Intelmq, that uses this threat intelligence for incident response. For effecting sharing of this threat intelligence, various standards are emerging which enables researchers, security professionals, academicians to share threat intelligence efficiently in a common machine digestible data format. In this paper we have done a survey of existing threat intelligence data sharing formats, we have categorized them based upon their significance to various user domains such as law enforcement agencies, Network administrators, researchers/ academicians.

## II. MOTIVATION

Attacks are dynamic and there is a need for a database repository that can provide data in a standard structured format. Attack data is required to detect unauthorized activities which are performed by hackers and attackers. The main purpose of network probe attack is the collection of data and to find the vulnerabilities (holes) in the network. The data includes IP address, e-mail address, host name, name of service, country address operating system application, etc., these are necessary for the attacker [4], [5]. The scam program, "which is used to collect the network data" is utilized by the attacker and to make its use in the primitive of another attacks. So, the sharing mechanism of captured automated attack data is necessary for further investigation.

## III. SURVEY ATTACK DATA CAPTURING SHARING INFRASTRUCTURE AVAILABLE

**Malware attribute enumeration and characterization (MAEC):** MAEC was firstly governed by MITRE and was published in 2010. It is represented in XML based structured language for describing high level malware mechanism, behaviors and malware actions [6], [7], [8]. Cybox, MMDEF, Open IOC, Yara are similar to MAEC and used in MAEC, whereas MAEC is used in STIX [9]. The various tools are used in MAEC are Anubis, ThreatTrack, ThreatExpert, Cuckoo Sandbox, Thug. The main target of Malware Attribute Enumeration and Characterization (MAEC) is to provide the transformation of malware research and response. The purpose of MAEC is to get rid of uncertainty and deception found in malware descriptions and to reduce dependency on signatures.

**Common Attack Pattern Enumeration and classification (CAPEC):** CAPEC was firstly governed by MITRE and was published in 2008. CAPEC is a language which is useful to define the list of common attack pattern. Cybox provides the information about the actions which are performed by the attacker regularly and is also helpful to mention the steps which involved in exploiting vulnerability [7] [8], [10]. CAPEC is basically a catalog in which attack patterns are mentioned and this is publically available. CAPEC is used by STIX, IODEF-SCI and Cybox is used by CAPEC.

**Cyber Observable Expression (Cybox):** Cyber Observable Expression was governed by MITRE and was published in 2012. Cybox is standardized XML based language which is used to represent observables in the operational domain for specification, capturing, characterization and communication of events or stately properties. It is helpful to express events such as file deletion, events, changes to registry keys values and communication via HTTP which would takes place at the time of attack [8], [11]. Cybox describes states and properties of IT aid in various ways. They are also used for performing incident response, detection of malware, event management etc. It uses pcap whereas it is used in STIX, CAPEC, and MAEC to represent observables. Cybox includes some indicators such as OS artifacts, APIs, X.509 certificates, network flows, network artifacts, files, SMS message, images, email messages. Various tools used in Cybox are Python-Cybox, Cybiet.

**Trusted Automated Exchange of Indicator Information (TAXII),:** Trusted automated exchange of indicator information was governed by MITRE. It is an XML based language. It provides threat information sharing between trusted entities [7], [8]. It is a transport protocol for STIX. TAXII was developed to achieve the simplicity and speed of sharing information across the organizations. HTTP protocol is used by TAXII for message transfer. TAXII is implemented in YETI and is written in Python [12], [13].

**Structured Threat Information Expression (STIX)** Structured Threat Information Expression was governed by MITRE and DHS and was published in 2012. It is XML based language and is expressive, flexible, and extensible. It is used to convey the potential cyber threat information. The exchange of data involves in indicators, adversary tactics, cyber observable, incidents, exploits [7], [8]. The main use cases of STIX are: analyzing threats, specifying indicators for threats, managing prevention and response activities, and sharing the threat information. MAEC, CVRF, OVAL, CAPEC, Snort signatures, CVSS, OpenIOC, TLP, CPE, CWE,CAPEC, YARA signatures CPE,TLP,OSVDB. Microsoft Interflow, CRITs, MANTIS, python-STIX are the tools for STIX.

According to **"Panos Kampanakis",** the most richest and prevalent options for sharing the attack data information are STIX and OpenIOC [7]. He also recommends that the best standard can be chosen by identifying the use cases according to the need of the data which will be shared and exchanged.

## IV. COMPARATIVE ANALYSIS OF ATTACK FEEDS OFFERED BY THEM

TABLE I.

| Sr. no | Feeds | Capturing technique | Data feeds | Commercial/non-commercial |
|---|---|---|---|---|
| 1. | CLEAN-MX Real-time Database | | URL,IP, Domain | |
| 2. | PhishTank Phish Archive | | URL | |
| 3. | Cyber Crime Tracker | | IP,URL | |
| 4. | Zeus Tracker | | IP,<br>Host name | |
| 5. | Team Cymru | Honeypots, crawlers, and leverage private data sharing agreements with partners. | Malware hash registry | Commercial/permission |
| 6. | Vx Vault | | URL,IP | Public |
| 7. | Malware domain list | | Domain, IP | |
| 8. | Palevo Tracker | | Domain, IP | |
| 9. | Malcode | | URL | Public |
| 10. | Malware Domainlist | | URL | Public |
| 11. | Abuse. CH Palevo Tracker | | Palevo command and control servers | Public |
| 12. | Abuse.<br>CH Feodo Tracker | | Feodo command and control servers,c & cs. | Public |
| 13. | ATLAS | | | Commercial |
| 14. | Spam<br>Haus Technology | | | Commercial |
| 15. | Sinkhole | | Virus,zeus,spyeye,,S D bot, and other malware infections. | Commercial |
| 16 | Dragon<br>SSH | | SSH Bruteforce attack | Public,non-commercial,permission for redistribution. |

## V. CHALLENGES IN DEVELOPMENT/AUTOMATED SHARING OF ATTACK DATA

The main challenge for the development of the attack data sharing is that, the data should be machine digestible [14], [5]. The information about the attack data should be human-readable, accurate, easy to understand, so that it could be used accurately for further investigation. For sharing the information of attack data some standards and platforms are needed. How to choose accurate standard is also the challenge. The accurate standard is chosen according to our need by identifying the use cases [15].

**Models for sharing attack data :** EU's -center of network and information security expertise - The European Union Agency for Network and Information Security (ENISA) has categorized the information sharing standards as under [14], [5]:

- Formats for low-level data
- Actionable observables
- Enumerations
- Scoring and measurement frameworks
- Reporting formats
- High level frameworks
- Transport and Serialization.

Information management tools categorized by ENISA as under:

- Automated distribution of data
- Supporting analysis
- General purpose log management
- Handling high level information.

As focus of our paper is reporting format of Information sharing, these are given as follows:

**A. Abuse Reporting Format (ARF)-** ARF is coordinated by Yakov Shafranovich and was published in 2005 [8]. ARF is text based developed for e-mail spam and is extension to MIME. Some e-mails contains spam reports, ARF allows the creation of these types of e-mail messages with spam samples attached. The report of the original spam message includes the source IP of the message, original message ID, the date of received message, and a message classification such as abuse, spam, virus, other, non-spam. This is superseded by MARF. Types of indicators used are spam reports and samples.

**B. Common vulnerability reporting framework (CVRF)-** CVRF is governed by ICASI and was published in May 2011. CVRF is XML-based language which is designed to provide a standard format for dissemination of security related information [7], [8]. Time reduction in problem resolution, vulnerability reports are automatically process and participation in a standard format with strong multivendor support. CVRF uses CWE and CVE which indicates product vulnerabilities. For this no tool is available publically, it is only used internally in vendor’s communications.

C. **Incident object description exchange format (IODEF)-** IODEF is governed by Managed incident lightweight exchange working group (MILE) and was published in 2007. It is XML based format which is designed to exchange the information between CSIRTS to provide the way to exchange the information of operational and statistic security incident [7], [8]. IODEF provides the compatibility with IDMEF and is extended to IODEF- SCI. Timing, incident description with confidence rating, network and OS artifacts, exploit and vulnerability references, contact information and incident history are the main indicators used by IODEF. The tools used are Collective Intelligence Framework (CIF) and Arcsight.

**D. IODEF for Structured Cyber security Information-** IODEF-SCI was governed by NICT and was published in 2014. It is XML based format which is the extended version of IODEF and is provide the extension for embedding structured information with in the document [8]. For embedded information, the other formats are used for the representation. It shows the relationships with CAPEC, CVE, CVRF, CCE, CWE, CPE, CVSS, CWSS, OVAL, XCCDF, CRE. The main symbols used are attack patterns, event reports platforms, vulnerabilities, weaknesses, scores, incident remediation description, verification checklists. It works on the IODEF SCI tools.

**E. Structured Threat Information Expression (STIX)-** STIX is governed by MITRE and DHS and was published in 2012. STIX is an artistic, flexible, and extensible XML based language that conveys potential cyber threat information. The exchanged information of TAXII is represented in the XML-based Structured Threat Information Expression (STIX) language [15]. STIX is an open source and is developed in the platform which is synergized. In this the data exchanged includes cyber observables, indicators, incidents, adversary tactics, exploits, and courses of action as well as cyber-attack campaigns and cyber-threat actors. STIX can bound on many other structured XML-based languages, like as Snort and Yara. The following are used by STIX are CYBOX, MAEC, OVAL, CAPEC, CWE, CVRF, CVSS, Snort signatures, OSVDB, TLP, CPE, CAPEC, YARA signatures OpenIOC [9]. The tools used by STIX are Microsoft Interflow, CRITs, Mantis, and python-STIX.

## VI. COMPARATIVE ANALYSIS OF THEM BASED UPON THE FEATURES OFFERED:

**A. TAXII-** TAXII (Trusted Automated exchange of Indicator Information) which is represented in XML format is the main system for transporting cyber threat information [16]. TAXII mainly have three models of sharing: Hub and scope, source and peer-to-peer. Services defined by TAXII: There are some services which are defined by TAXII, for different sharing models:

- Inbox: Pushed contents are received in this service.
- Poll: This is the service to request content.
- Collection Management**:** This service may be defined as, “in which the data is read about and the request is accepted to data collections”.
- Discovery: Discovery may be defined as, which provides the information of supported services and also tells that how to develop connections with them.

Cyber Observable Expression (CybOX) and Structured Threat Information Expression (STIX) both are the open source formats which are helpful to exchange the information of cyber threats in automated form. This is mainly helping in representation of cyber-threat information in a standardized format. They are basically standards that software can use. When STIX and TAXII both are used in combination then it is more helpful to anybody to share the threat information in easiest way with people.

*STIX*- If any organization wants flexibility to share the threat information then the services provided by TAXI are helpful to share the information automatically.

Closely working communities STIX and TAXII ensure that full stack for sharing threat intelligence are provided by them [25].

**B. CYBOX-** CybOX (Cyber Observable expression) is a language in which the events are described. The described events contain the stately properties. Hence the cyber domain is used to observe these properties of events. [7], [11]. For this purpose Cybox is beneficial for STIX, used to indicate patterns, parameters for course of action and infrastructure descriptions. The communities and organizations which are working on STIX and Cybox, "are providing the information that Cybox is beneficial when used independently". Cybox is also beneficial if there is requirement of use cases of STIX then it is helpful to support the use cases also.

**C. MAEC**- MAEC (Malware Attribute Enumeration and Classification) is a language which describes the malware behavior and provides response to malware analysis [6], [17].

If the three specifications are combined together then they can interact and support individually as well as have usage in combination.

**D. CAPEC-** Initiative Started by Department of Homeland and security "Common Attack Pattern Enumeration and Classification" (CAPEC™) with the purpose to develop the standard system to identify, collection, refining, and sharing of attack patterns among the software communities is now used by the STIX for characterization of framework of (TTP) Techniques and Procedures Attack Patterns with the help of CAPEC schema extension [10].

**E. IODEF-** The Incident Object Description Format (IODEF) is an Internet Engineering Task Force (IETF) standard which aims to exchanging incident information, [3], [4]. But if we deeply see there is no exactly relationship exists between IODEF and STIX, no doubt it is possible to hold IODEF within STIX for representation of incident information. But some risks are involved such as loses the richness and alignment of architecture which is provide by STIX Incident structure.

**F. OpenIOC-** The STIX Indicator's test phenomenon is a good alternative for providing a signature indicator in something other than CybOX [7], [9]. Open Indicators of Compromise, And default extensions that support the test phenomenon is Open Vulnerability and Assessment Language (OVAL), YARA rules and SNORT rules.

**G. CIQ-** The OASIS Customer Information Quality (CIQ) is a language used to represent individuals and organizations information. As the STIX Identity structure which is usesd as an extension mechanism for identification of information used to characterize malicious actors, victims and intelligence sources. The STIX-provided an advance extension CIQ.

**H. VERIS-** The Vocabulary for Event Recording and Incident Sharing (VERIS) - framework used to collect risk management information from incident security and also helps to provide the common language for describing their effects in a structured manner. VERIS basically intended to use for post-incident strategic trend analysis and risk management [7], [9] [21].

On the other hand if STIX is used in extensive threat intelligence framework then it is helpful to capture information of security incidents and is also helpful to provide the information of effects of the captured security incidents [18]. Some organizations such as STIX community- Verizon and members of VERIS team etc. are frequently working on all the above explained formats. Due to some limitations in STIX format, "as the information in this format is not refined", they are switching on the VERIS format. VERIS format is helpful to improve the data. It provides the refined data of the STIX incident schema. VERIS input has an important role in good portion of STIX Incident schema derivation.

### I. CSV (Comma separated Value) Format

It is one of the sharing formats for cyber threat information in real time which is also supported by CFM; it is a file format which helps to store data which looks like a text file. The data is characterized and stored in manner of one record for each line and comma is used for separation of each and every field. CSV is a format which is easy to understand by human can be easily edited manually, easy to implement, can be easily parsed, and provides straight forward information schema [22].

### J. JSON Format

JSON is the format which is text. It is used to exchanging the data between browser and the server. JSON, which is received from server can be converted in JavaScript. But this format is text only it cannot gives the output as comma separated values. Hence this format is very hard for read for humans [23].

**Conclusion-** It is concluded with the thoroughly study of literature survey that lots of sharing standards are there with special features. To selecting a standard sharing format is very crucial in society as it should be easily understandable, very informative, less time consuming and should be in the machine readable form. So the standard structured format that we have selected for automated sharing of information on attack data is the CSV (Comma Separated Values) format. As all the information about targeted attacks will automatically generated using the honeypots and ELK technology and shared with the society in a CSV format [22].

## REFERENCES

[1] Dhanashri Ashok Bhosale on Comparative Study and Analysis of Network Intrusion Detection Tools in 2015.

[2] Munish Sharma, Tejinder Kaur on A Study on Network Intrusion Detection Based on Proactive Mechanism, 2014

[3] Snehal B Rase1 , Pranjali Deshmukh on Summarization of Honeypot- A Evolutionary Technology for Securing Data over Network, And Comparison with some Security Techniques in 2015

[4] The MITRE Corporation on Cyber Information-Sharing Models: An Overview in October 2012.

[5] Cristin Goodwin, J. Paul Nicholas. A framework for cyber security information sharing and risk reduction in 2015.

[6] Vijay Varadharajan. On Malware Characterization and Attack Classification in Information and Networked Systems Security Research Macquarie University, Australia in 2013.

[7] Panos Kampanakis. Security automation and threat information- Sharing options, co-publish by the IEEE computer and reliability
societies in September/ October 2014.

[8] European Union Agency for Network and Information Security (ENISA) on Standards and tools for exchange and processing of actionable information in November 2014

[9] “Open Indicators of Compromise (OpenIOC),” Mandiant, 2013; http://openioc.org.

[10] The MITRE Corporation on Common Attack Pattern Enumeration and Classification — CAPEC™ A Community Knowledge Resource for Building Secure Software.

[11] European Union Agency for Network and Information Security (ENISA) on Detect, share, Protect Solutions for Improving Threat Data Exchange among CERTs in October 2013.

[12] Rafiqul Islam, Ronghua Tian, Lynn Batten and Steve Versteeg on Classification of Malware Based on String and Function Feature Selection in the workshop of IEEE, Second Cybercrime and Trustworthy Computing in 2010, pp. 9-17.

[13] Kai Huang, Yanfang Ye, and Qinshan Jiang. Ismcs: an intelligent instruction sequence based malware categorization system. In ASID'09: Proceedings of the 3rd international conference on Anti-Counterfeiting, security, and identification in communication, pages 509–512, Piscataway, NJ, USA, 2009. IEEE Press.

[14] Jessica Steinberger, Anna Sperottoz, Mario Gollingy and Haral Baier. How to Exchange Security Events? Overview and Evaluation of Formats and Protocols in Biometrics and Internet Security Research. Group University of Applied Sciences Darmstadt, Darmstadt, Germany2015.

[15] Sean Barnum on Standardizing Cyber Threat Intelligence Information with the Structured Threat Information expression (STIX™) in Feb 20, 2014.

[16]Julie Connolly, Mark Davidson, Matt Richard, Clem Skorupka. The MITRE Corporation on The Trusted Automated exchange of Indicator Information (TAXII™) in August 2012.

[17] Desiree Beck, Penny Chase, Ivan Kirillov, MITRE on the MAEC™ language V4.0.1 detailed examples in Feb24,2014.

[18] Mike Schiffman, Cisco Systems on The Common Vulnerability Reporting Framework An Internet Consortium for Advancement of Security on the Internet (ICASI) Whitepaper in 2011.

[19] Navroop Kaur, Amit Bindal on Complete Dynamic Malware Analysis in 2016.

[20] Gaurav Kumar on Impact of Agile Methodology on Software Development Process in 2012.

[21]"VERIS," Verizon; http://veriscommunity.net.

[22] Richard Zuech, Taghi M. Khoshgoftaar, Naeem Seliya, Maryam M. Najafabadi, and Clifford Kemp on A New Intrusion Detection Benchmarking in Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference in 2015, pages 252-255.

[23] B. Navya Rupa, G. Krishna Mohan, J. Satish Babu and Tai-hoon Kim on Test Report Generation Using JSON in International Journal of Software Engineering and Its Applications in 2015, Vol. 9, No. 6, pp. 63-70.