Lets Play the Game of Botnet

Deepti Aggarwal IGDTUW ,Vanshika Madan
The NorthCap University,Utham Kuma

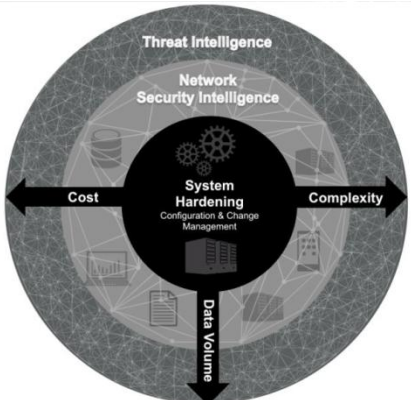
E-mail deepti937@gmail.com

Abstract— **With the rapid increase in usage of network based applications, the concern of securing the content of these networks is becoming an issue. Organizations now-a-days are investing more and more money to secure their data from the attackers. On the other hand, the attackers are getting stronger day by day. Hence it's necessary to bring an intelligence model which analyses and reports in advance of the possible cyber-attacks on the network .In this paper we'll be developing a cyber-threat intelligence (CTI) model which will be used to capture the loop holes of the network causing the intrusions. Intrusions in an information system are the activities that violate the security policy of the system. So an intrusion detection system needs to be developed which can monitor network for any harmful activities and generate results to the management authority. CTI model aims at developing an intelligent Intrusion Detection System (IDS) and Intrusion Prevention System (IPS) using Genetic Algorithm and Fuzzy Logic of machine learning.**

***Keywords*— Include at least 4 keywords or phrases**

I. INTRODUCTION

Threat intelligence is getting organized, analysed and refined information about potential or current attacks that threaten an organization.The primary purpose of threat intelligence is helping understand the risks of threats, such as advanced persistent threats (APTs), Ransomware and exploits. It is not only about what intelligence is collected, but also how it is analysed and used.



Threat intelligence includes in-depth information about specific threats to help an organization protect itself from the types of attacks that could do them the most damage.
Threat intelligence of an organisation depends on three major factors:-

1. **COST**:
   While choosing a threat intelligence system an organization always checks whether the system is in our budget or not . A company with smaller turnover will always demand for a threat intelligence system of lesser cost .Whereas a bigger company would find itself in a position to afford an intelligence system of higher cost.
   Therefore costing of a threat intelligence system must be considered as an important factor while choosing or creating an intelligence system as it varies from one organization to other.
2. **Data Volume**:
   Spending more on threat intelligence programs is a good start, but it isn't enough.The threat intelligence model also depends on the volume of data that is if there is low data utilization we need not reserve large amount of resources for it. However the model should be such that it should cater not only to present demands but should also be scalable.
3. **Complexity:**

   According to classical studies, a pro-active surveillance based cyber-threat intelligence model is formed at two levels, strategic and tactical, respectively.
   Strategic model is about planning a base to handle an attack. Invaders social , political, technical, economic and cultural motives is collected to plan the defence effectively. This intelligence can be made available with trends and patterns seen through tactical intelligence. Attackers have a tendency to change their attack pattern but their motive remains same, hence strategic intelligence is an important term to deal while building a CTI model.
   The tactical model is formed by observing the attack for threat indicators. This observation is generally documented via Indicator of Compromise (IOCs). IOC mostly document indicators such as URL redirections, system rebooting IP Addresses (IPv4 & IPv6), File Hashes, Email addresses. Therefore, tactical intelligence will cover defence and monitoring, security information and event

management (SIEM) and simultaneously feed input to build strategic intelligence.

The paper is divided into four sections. Section I, Introduction introduces about two levels at which cyber threat intelligence model (CTIM) can be studied. Section II describes the standard expectations from a CTIM. Section III covers various models present in market and used among organizations. Section IV tells about the loop holes of existing models. Section V gives the proposed solution "Cyber Threat Intelligence" model and how it is better from other CTI models.

II. EXPECTATIONS FROM A CIT MODEL

A cyber-threat intelligence model is expected to find most suspicious activities by analysing and reporting all the possible cyber-attacks in advance. Threat intelligence is evidence based, knowledge gaining, about an existing or emerging hazard that can be used to make decisions regarding the subject's response to that menace or hazard.

Inputs to a CTIM are threat feeds and these are categorised as: internal threat feeds, external threat feeds and community threat feeds.

a. Internal threat feeds

Sensitive/system/server/networks**,** Personal data systems**,** Sensitive user list**,** Web/externally accessible platforms, Mapping of IP address to office locations.

b. External threat feeds

Compromised web sites/ URLs, Botnet memberships/spam sources, Phishing/attack email subjects, Physical location, Countries/network/location likely to initiate the attack

c. Community threat feeds

Pattern of attacks in one area that can be monitored for re-occurrence in others, Inter-domain information flows, traffic or connections,Known system vulnerabilities and in-the-wild exploits.

Along with input feeds, CTIM also needs to persistently store and analyse various logs like infrastructure logs, application logs and technology configuration data. All of this input data collectively serve for future research and analysis process.As per the SANS report Tools and Standards for Cyber Threat Intelligence Projects, 2013 following requirements from a CTIM are identified:

R1	Capability to Import/Export indicator details to/from other systems in a standard format.
R2	Capability to Import/Export structured incident data to/from other systems in a standard format.
R3	Capability to Query, Import, Export and Manage CTI data through a user interface.
R4	Capability to enforce data sharing based on an attribute attached to CTI data.
R5	Capability to automate the import and export of CTI data.
R6	Capability to provide authentication and confidentiality when sharing data.
R7	Capability to export data that can be used in detective and preventive controls.
R8	Capability to select data for export based on creation dates of CTI data.
R9	Capability to measure the efficacy of CTI feeds.

Our current challenge is not only to make a CTI model but is also, to improve the efficiency of existing models by considering two of the requirements mentioned above R2 and R9 which are to import/export structured incident data from/to other systems in a standard format and capable of measuring the efficiency of CTI feeds respectively. These two requirements are least trailed by the existing model which we have studied in this paper.

III. VERIS EXISTING THREAT MODELS

As we have discussed earlier in this paper, CTI Models are based on strategic and tactical intelligence.

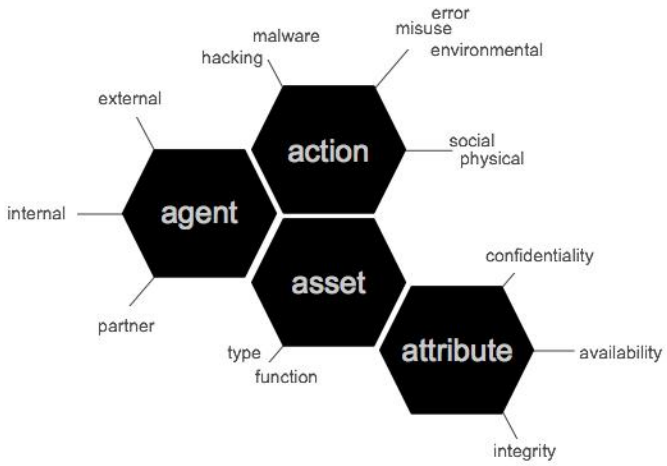
A. The strategic based intelligence model includes Verizon's VERIS. The tactical intelligence based model includes Mandiant'sOpenIOC Framework, Alient Vault's Open Threat Exchange, Collective Intelligence Framework, Threat Connect, Hailataxii , IBM's X-force. Mitre Standards for collecting cyber intelligence- CybOX for cyber observables, STIX for defining and TAXII for sharing CTI among multiple threat sharing communities

The Vocabulary for Event Recording and Incident Sharing (VERIS) is an open source threat model designed by Verizon to provide a common language for describing security incidents in a structured and repeatable manner. It is a strategic based model which works on JSON language.VERIS deals with "lack of quality information" which is considered to be one of the most critical and persistent challenges in the security industry. VERIS collects useful incident related information from the network and share that information anonymously and responsibly-with others. VERIS schema defines variables at its basic level for descriptive enumeration of security breaches. VERIS model is divided majorly into 5 sections each of which is designed to capture a different aspect of the incident narrative:

1. Incident Tracking
2. Victim Demographics
3. Incident Description
4. Discovery & Response
5. Impact Assessment

It further maps each security event with 4 A's i.e.

1. Agent- Whose actions affected the asset
2. Action-What actions affected the asset
3. Asset- Which assets were affected
4. Attribute-How the asset were affected



Additionally, Verizon annually releases Data Breach Investigations Report (DBIR) for community reported breaches. Support for VERIS can easily be found on github since it is an open source model. The model fulfils requirement R2 as identified by SANS report. The only with VERIS is its limited support for gathering tactical intelligence.

B. Open IOC Framework

Open Indicators of Compromise (Open IOC) is a tactical based intelligence gathering threat model. It is a full tie open source model of version 1.1 working under Mandiant. Open IOC is designed to fill a void that currently exists for organizations that want to share threat information both internally and externally in a machine-digestible format. Open IOC is an extensible XML schema that enables you to describe the technical characteristics that identify a known threat, an attacker's methodology, or other evidence of compromise. Open IOC offers your organization the option of using MANDIANT's field tested Indicators of Compromise. These indicators describe over 500 facets of environments that can be used to track down advanced attackers. The tangible framework meets requirement R1 and R2 only.

C. CybOX

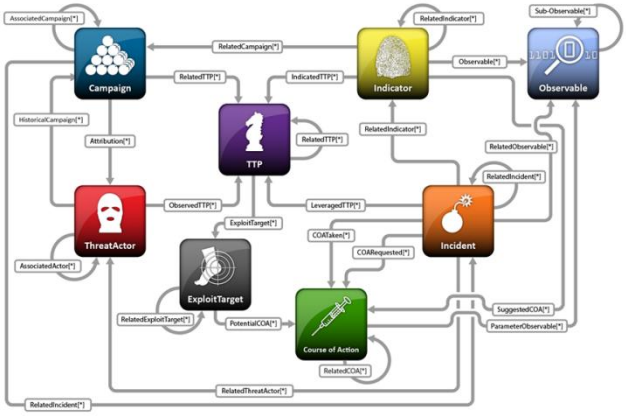
Cyber Observable eXpression (CybOX) is a first tactical based threat intelligence model developed by Mitre Corporation. CybOX is a standardized language for encoding and communicating high- reliability information about cyber observables. This model is used for representation of cyber observables in the form of threat intelligence, malware characterization, security operations, indicator sharing etc.By specifying a common structured schematic mechanism for these cyber observables, the intent of this model is to enable the potential for detailed automatable sharing, mapping, detection and analysis heuristics.CybOX uses 70+ defined objects including X.509 Certificate, Linux Package, Email message, Domain Name, DNS Record etc. that can be used to define measurable events.

CybOX has also got a huge support for APIs and Tools, for example OpenIOC-to-CybOX is a python utility to import Mandiant'sOpenIOC format into CybOX and Email-to-CybOX is another python utility to convert email into CybOX Observables documents.

This model follows the requirement of R1 and R2.

D. STIX

Structured Threat Information Expression (STIX) is a structured language for describing cyber threat information so that it can be shared, stored, and analysed in a consistent manner.Core use for STIX is the sharing of cyber threat information like Observables, IoCs, Exploit Targets, Cyber Attack Campaigns etc. within an organization and with outside partners or communities that benefit from the information. It can be used manually or programmatically. Manual use requires an XML editor, but no additional tools. Programmatic use requires Python and Java bindings, Python APIs and utilities to convert Mandiat'sOpenIOC Framework format into STIX IoCs. Requirement R1 and R2 are followed by the STIX model. Fig. 3 shows the underlying architecture of STIX.



E. TAXII

Trusted Automated eXchange of Indicator Information-TAXII model, based on tactical intelligence is used for automated exchange of cyber threat information using STIX for expressing cyber threat information. Organizations can use TAXII services and message exchanges in order to share security event information with one another so that evolving cyber threats and attacks can be sensed and alleviated more quickly than with existing non-information sharing technologies. TAXII is tightly coupled to other initiatives such as STIX and CybOX.The inter-relationship of these initiatives can be explained in a way that STIX uses language such as (but not limited to) CybOX to represent cyber security event information, and it also serves as the transport mechanism for

STIX information. Source-Subscriber, Peer-to-Peer, Hub -and-Spoke are the networks supported by the TAXII to import/export threat feeds. This brings outs the confidentiality and authentication of source and destination which is covered as requirement R6 by this model.

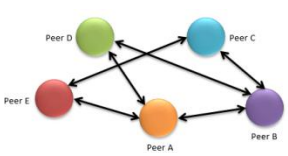


Fig. Peer-to-Peer

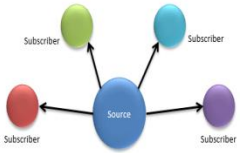


Fig. Source-Subscriber

Fig. Hub -and-Spoke

F. Collective Intelligence Framework (CIF)

CIF model based on tactical threat intelligence was developed by CSIRT Gadgets. This model mainly aims at collecting threat feeds in the form of IP addresses, URLs and domains to combine malicious information for identification, detection and mitigation. It includes concepts of severity and confidence as well as privilege which allow you to provide feeds of high-confidence public data to some systems while still allowing investigators to query private, unconfirmed data. CIF comes preconfigured with data sources like Shadowserver, Spamhaus, ThreatExpert etc. allowing it to run queries either through web browser, CLI client or through API fulfilling requirement R3. Data warehouse supported by CIF are JSON and ElasticSearch leading to fulfilment of maximum number of requirements R1, R3, R4, R5, R6, R7, R8 and R9.

G. Open Threat Exchange (OTX)

Alient Vault's OTX is an open source tactical intelligence based model which collects IoCs - IP Address, Domains, File Hashes, Hostnames, CIDR Rules, CVE number etc. fulfilling requirement R1 and thereby feeds this intelligence to define strategic intelligence in the form of pulses. Pulses provides with the summarized threat view i.e. target information, details of environment exposure etc.Like TAXII, OTX also automates mechanism for sharing CTI, this fulfilling requirement R5. OTX function on pulses which in turn are IoCs. The only limitation OTX possess is its anonymous input of threat feeds i.e. authentication of threat feed owner is not promised by OTX.

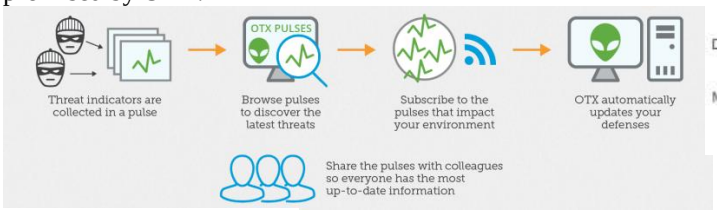
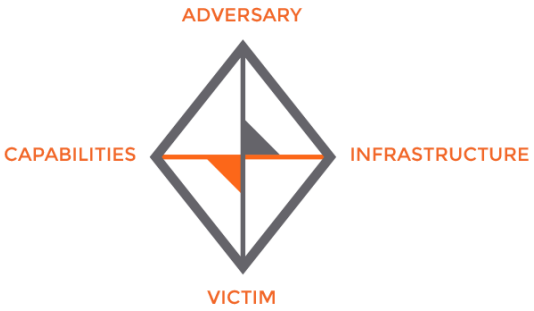


Fig. Flow of pulses in OTX

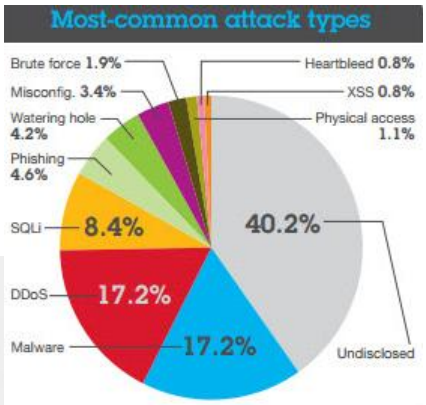
H. Threat Connect

Threat Connect, launched officially at Black Hat, 2013 is an open source (limited functionality) as well as closed source (full functionality) tactical intelligence based model. This model is at once simple and complex, informal and formal, useful for analysis of both insider and external threats fulfilling requirement R1 and R2 partially. It uses Diamond Model for Intrusion Analysis for cyber threat information and Kill Chain for building strategic intelligence. Limitation of Threat Connect could be its restriction to export and import indicator details with limited open source communities thereby ensuring confidentiality and authentication when sharing data i.e. R6.



I. X-Force

IBM's X-Force, launched in 2015, is tactical based model that's collecting data since 1990. It provides an updated list of potentially malicious IP addresses and URLs. The list identifies any undesirable activity in your network environment before it threatens the stability of your network. X-Force uses a series of international data centers to collect tens of thousands of malware samples, analyze web pages and URLs, and run analysis to categorize IP address information. IBM X-force in Feb, 2016 released Fig. 9 which reports most common attack types found in its malware research.



IV. SOLUTION PROPOSED

Till now, we got to know that cyber-threat intelligence model should be able to find most suspicious activity. In the same regard, we wish to invent a cyber-threat intelligence model to capture real time malicious traffic, process using advanced machine learning algorithms used for security and thus producing a product which is a combination of intelligence IDS and IPS.

Our solution aims at developing a model which can

-detect all the potential cyber threats

-prevent the occurrence of such attacks by deploying adequate counter measures

-respond in a precise way by understanding the motives, capabilities and objectives of threat actors.

-analyse and protect the high-value information assets.

-share all the gathered cyber threat information with industry peers to determine whether there are other threats and TTPs associated with the campaigns.

The model is developed in a combination of digital forensics, malware analysis and machine learning which is implemented on ELK stack.

Digital forensics: Digital Forensics can be defined as the use of scientifically derived and proven methods, towards the prevention,collection,validation,identification,analysis,interpretation,documentation and presentation of digital evidence, derived from digital sources, for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions.

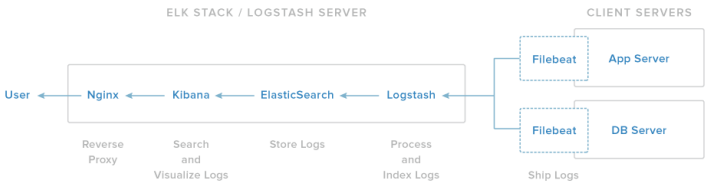
Malware analysis: Malware, also known as malicious code, is a common tool that attackers use to breach computer networks today, causing damage and disruption, and often requiring extensive recovery efforts.

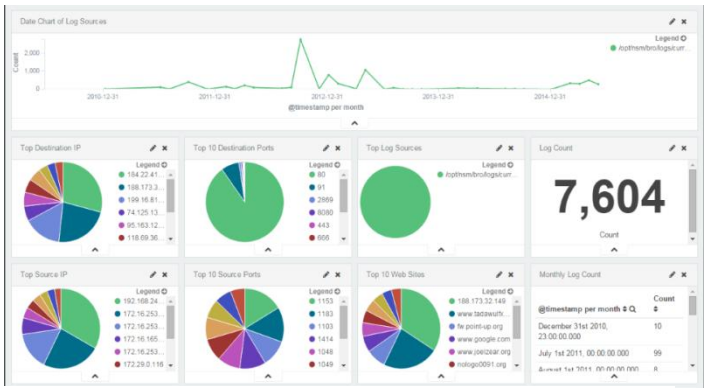
Machine learning: Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data.

ELK stack: ELK stands for ElasticSearch, Logstash and Kibana.

Our ELK stack setup has four main components:

1. Logstash: it is a server that processes incoming logs.
2. Elasticsearch: Stores and indexes all of the logs
3. Kibana: Web interface for searching and visualizing logs, which will be proxied through Nginx
4. Filebeat: Installed on client servers that will send their logs to Logstash, Filebeat serves as a log shipping agent that utilizes the lumberjack networking protocol to communicate with Logstash.





Since it's a model making, we have well defined inputs, outputs and a hypothesis with mathematical validation. We try to achieve a threshold of 65% with below stated technology stack.

INPUT:

For input data collection, Dionaea, high interaction honeypot is setup and corresponding data collection server, MySQL, is setup on operating system level virtualized machine.

Various Input to be collected include:

- Obfuscated code
- Network data is collected through TCP/UDP in the form of:
    - Exploit kit
    - APTs
    - Ransomware
- Packed executable are also feed in the form of files

Once input data is collected,

- Obfuscated code is converted into a structured format.
- Network data is collected using a script written in WireShark to transport in backend MySQL server.
- Packed executable are directly run.

CTI MODEL:

1. Input is fed to unsupervised machine learning algorithms – for e.g. Clustering.
2. Once clusters are formed, meta-data is extracted.
3. Feature selection is performed on 2.

4. Iterative supervised learning algorithms (for e.g. Decision trees, naïve bayes etc.) are fed on 3. Depending upon the best results, corresponding algorithms and results are extracted.
5. Note: We are considering Genetic Algorithms for APTs, Fuzzy + Neural for Ransomware. But if the data collection goes beyond limit, we may shift to Neuro Classification.

OUTPUT:

- Intelligent data, analyses and results.
- Intersections

For e.g. A = Locky
B = Cryptowall
C = Stuxnet
Intersections may be (A ∩ B ∩ C) or (A ∩ (B ∪ C)) targeted on windows, targeted through windows drivers and zero days.
Above example shows that after collecting intelligent data about the three attacks mentioned we'll get some common attributes which will help us to identify the common root cause and nature of such attacks.

MATHEMATICAL VALIDATION:

- Prediction for incoming attack comes from Probabilistic Modelling.
- Attack Patterns comes from Vector Algebra, Transformations and Co-ordinates.
- In the proposed hypothesis of this model, we take , three dimensional vectors where 'x' represents payload, 'y' represents meta data, 'z' represents attack pattern.
- The overall function goes by following

$Z = \sum(x,y,z)\ nk{=}1$

- The models' efficiency is computed by primarily computing
- Vulnerability estimation rate
- Statistical modelling of propagation of an attack
- Variations
- Time frequency analysis

FUTURE SCOPE

- Such a model if trained and deployed, may potentially write its own zero-days eliminating the need of pen-testers and security auditing teams.

Timeline.
During the course of three months development and research, following shall be our estimation for work distribution:
Month 1:

- High Interaction Honeypot setup
- ELK Stack distribution for threat intelligence data gathering, visualization and reporting Attacks on TOR and I2P
- Integrate snort IDS and honeypot.

Month 2:

- Writing Proof of Concept and Prototyping
- Scaling to convert into big data platform for detecting real time attacks

Month 3:

- Artificial based automation cyber threat intelligence is ready
- Increasing of Efficiency

EFFECTIVENESS OF THREAT FEEDS AND WHY DOES IT FAILS

V. CONCLUSIONS

THE PAPER ATTEMPTS TO COVER THE THREAT INTELLIGENCE MODELS PROMULGATED FROM 2009 TILL 2016. HOWEVER, THE PROGRESS IN INDUSTRY IS QUITE RAPID. THEREFORE, FURTHER ADDITION TO THE RESEARCH IN SUCH REGARD SHALL BE CONTINUED. FROM THE TABULAR SUMMARISED DATA, CLEAR ABSENCE OF A MODEL FOLLOWING REQUIREMENT R2 AND R9 IS EVIDENT. UNTIL UNLESS EFFICACY OF THREAT FEEDS IS COMPUTED, THERE WOULD BE LACK OF STANDARD INPUT TO A TIM. SOLUTION PROPOSED TO THE PROBLEM GIVES US THE ASSURANCE OF EFFECTIVE THREAT DETECTION AND PREVENTION MECHANISM

REFERENCES

[1] Dog, Spike E., et al. "Strategic Cyber Threat Intelligence Sharing: A Case Study of IDS Logs." Computer Communication and Networks (ICCCN), 2016 25th International Conference on. IEEE, 2016.
[2] Tools and Standards for Cyber Threat Intelligence Projects, October 14th 2013, SANS Institute InfoSec Reading Room, https://www.sans.org/readingroom/whitepapers/warfare/toolsstandards-cyber-threatintelligence-projects-34375
[3] https://github.com/vz-risk/veris (28/11/2016 19:20:16
[4] http://veriscommunity.net/howto.html (28/11/2016 19:24:16)
[5] http://openioc.org/ (28/11/2016 19:34:27)
[6] https://github.com/mandiant/ioc_writer (28/11/2016 20:04:14)

[7] http://www.verizonenterprise.com/verizon-insightslab/dbir/ (28/11/2016 20:31:18)
[8] https://github.com/mandiant/OpenIOC_1.1 (28/11/2016 21:56:28)
[10] Standardizing Cyber Threat Intelligence Information with the Structured Threat Information eXpression (STIX™), February 20, 2014, Mitre Co., https://www.standardscoordination.org/sites/default/file s/docs/STIX_ Whitepaper_v1.1.pdf
[11] https://taxiiproject.github.io/ (02/12/2016 15:12:19)
[12] https://securityintelligence.com/how-stix-taxii-andcybox-can-helpwith-standardizing-threat-information/ (02/12/2016 20:53:45)
[13] Alient Vault Threat Exchange, 2016, http://billows.com.tw/download/dm/AlienVault-OpenThreatExchange.pdf
[14] Caltagirone, Sergio, Andrew Pendergast, and Christopher Betz. "The diamond model of intrusion analysis" Center for Cyber Intelligence Analysis and Threat Research Hanover Md, 2013.
[15] Obrst, Leo, Penny Chase, and Richard Markeloff. "Developing an Ontology of the Cyber Security Domain." STIDS. 2012.
VerIS- a Framework for Gathering Risk Management Informa8on from Security Incidents, Wade Baker Alex Hutton Chris Porter, Risk Intelligence Verizon Cybertrust Security,
[9] http://cyboxproject.github.io/about/ (02/12/2016 14:20:20)
http://www.securitymetrics.org/attachments/Metricon - 4.5-BakerHutton-VERIS.pdf
[16] https://securityintelligence.com/a-gentle-introductionto-the-x-forceexchange-api/ (5/12/2016 18:59:12)
[17] Who's Using Cyberthreat Intelligence and How?, SANS Institute InfoSec Reading Room, February 2015, https://www.sans.org/readingroom/whitepapers/analyst /cyberthreat-intelligence-how-35767
[18] Osako, Takeshi, Tomoyoshi Suzuki, and Yoichi Iwata. "Proactive Defense Model Based on Cyber Threat Analysis." FUJITSU Sci. Tech. J 52.3 (2016): 72-77.
[19] Gartner. Definition: Threat Intelligence (EB/OL). https://www. gartner.com/doc/2487216/definitionthreat-intelligence,2013-5-16.
[20] http://csirtgadgets.org/collective-intelligenceframework (03/12/2016 14:33:12)
[21] https://github.com/csirtgadgets/massive-octo-spice (03/12/2016 14:37:12)