Tallinn University of Technology
Faculty of Information Technology
Institute of Computer Science
IVCM09/09 – Cyber Security

Rossella Mattioli

Information exchange framework for cyber security incidents

Master's Thesis

2012

Supervisors:

Prof. Ahto Buldas
Tallinn University of Technology

Jose Nazario, Ph.D.
Senior Manager of Security Research
Arbor Networks

Tallinn, 28th of May 2012

Hereby I declare that this Master's Thesis, my investigations and achievement, submitted for the Master's Degree at Tallinn University of Technology has not been submitted for any academic degree or examination at any other university.

Rossella Mattioli

107974IVCM

The Thesis conforms to the requirements of Master's thesis.
Write your supervisor's name here ..... (Signature and date)
Chairman of the Defense Committee ..... (Signature and date)

Annotation

The purpose of this thesis is to define the initial baseline for a wider research focused on providing a comprehensive cyber security incident information exchange framework for response teams. The present document aims to investigate the feasibility of this approach and has identified the following preparatory steps. First, the analysis of a corpus of publicly available data feeds will be deployed in order to understand which kind of contents are accessible to the use of response teams. Then it will be proposed an initial taxonomy of incident indicators in which the analyzed data will be recollected. This will allow to have a common exchanging platform where to start categorizing the content. Consequently a first implementation via a widely recognized standard and through the use of other available specifications and formats will be proposed. In the end a feasible scenario of implementation covering the potential requirements of a system that could be deployed will be presented including an example based on the data of a widely recognized incident.

Infovahetuse raamistik küberturbe intsidentidele

Töö eesmärk on luua alus laiemale uurimistööle, mis käsitleb ühtse keele loomist turvameeskondadele küberturbe intsidentide kohta käiva informatsiooni edastuseks. Töös uuritakse sellise lähenemise teostatavust ja astutakse mõned ettevalmistavad sammud. Esmalt analüüsitakse avalikku informatsiooni turvaintsidentide kirjeldustest, et saada ettekujutus turvameeskondadele ligipääsetavast informatsioonist ja tekkivast keelekorpusest. Seejärel pakutakse välja esialgne indikaatorite süsteem, mille põhjal oleks võimalik koguda lähteandmeid ja alustada nende sisu kategoriseerimist. Seejärel pakutakse välja esimene olemasolevatel standarditel põhinev realisatsioon ja katsetatakse seda ühe laialt tuntud turvaintsidendi kohta teada olevatel andmetel.

To Sara, the little dragonfly,
and
to my nephew and two nieces who are on the way to this universe.

“In cases of major discrepancy it is always reality that's got it wrong.

And remember, DON'T PANIC.”

RFC 1118 - The Hitchhikers Guide to the Internet

Table of Contents

Index of Figures

Index of Appendices

1 Introduction

The more we build the Information Society, the more every day we expose people using IT services to a crescent volume of cyber security incidents [i]. Every day thousands of information security events happen all over the world due to the ubiquitous nature of the Internet, cyber criminals are no more bounded by the national borders. Every day in almost every country[ii] Incident Response Teams (IRT), Computer Emergency Response Teams (CERTs) or Computer Security Incident Response Teams (CSIRTs) collaborate and face these events in a heterogeneous way. This is not only due to the different kind of legal frameworks and constituencies but also to the different standards and technical solutions they use. The level of information available to this kind of entities is as wide and heterogeneous as the Internet is. The aim of this thesis is to map the current situation of data available to response teams and define a baseline where to start leveraging security events exchange. At the moment there is no solution which comprehensively answers all these needs so this thesis represents the start of research focused on filling this gap. The goal is to provide a comprehensive framework and application to exchange and face security incidents in a timely and consistent manner.

1.1 Research Goal

The purpose of this thesis is define the initial proposal for a wider research that will be focused on providing an end to end solution primarily for response teams (IRT, CERT, CSIRT and abuse teams) and also all entities involved in cyber security incident exchange. End to end means that it will cover the process from the entry of incident data into the system of a response team, its categorization and correlation to the follow-up output and post-mortem analysis. This research aims to propose a solution that can pave the road for a open discussion on the future developments of incident exchange among all actors involved in cyber security scenarios. The goal is to make use of available data sets, models and formats and create a scalable framework that will answer daily operational requirements of response teams and enhance collaboration among them and all the entities involved.

1.2 Framework and Application Outline

Security events exchange among computer security response teams is not just a matter of technology. Receiving an input regarding an anomaly in the system, whether it is something coming from an intrusion detection system (IDS), a clearing house, an internal user or an

external source, can trigger not only a technological response but raise also several questions regarding different aspects, such as legal, organizational and financial issues. These non-technological aspects are becoming more important day after day, but as it will become clear in this research, before addressing them there is the need to align the information input and exchange baseline. The main idea is to create a comprehensive and flexible information exchange base. Such a base will give the response teams a global view of all data available and will enable then to face legal, organizational and economic issues in a more consistent way. As it will be shown, the information are already present and how to define a way to categorize and correlate them in an organic way should be the first area of research.

The main idea behind this research is to provide computer response teams with a global framework that would allow them to have an extensive view of all the threats that could harm their networks or that originates from their networks. This view is intended to be an aggregation and representation of all the data sources available that will be presented in the following chapters. As it will be shown the problem is not about gathering information but categorizing the data and displaying relevant information to the users. Assuming that a security event is an agglomeration of indicators, it can be useful to define a system that identifies these basic components and correlates them. The different feeds that are available to a response team should be considered in a holistic way and not as separate entities. Most of the internal or external feeds available to a response teams can have several points in common and try to standardize them can be useful to have a complete overview of a specific incident. Considering the data flows in their point of connections and not as separate entities can gain a competitive advantage in facing even the most dangerous threats. This approach enables a more organic evaluation or response to a threat and can also provide a standardized framework not only for input but also for output and post-mortem analysis.

During the presentation of the current situation it will become clear that there is no need to create a new standards or specifications. The current need is to develop a system that enables the aggregation of all basic components that are common in various feeds independently from their source and make use of those specifications that can align them and pave the road for future implementations. In applying this approach the data should be parsed using a standard that can be easily deployed and set the ground for future developments. For the response team the parsing operation should be transparent and easily scalable. In this case Incident Object Description Exchange Format (IODEF)[iii] data model became the ideal candidate. Due to the

definition of all the different classes and the granularity of its attributes it covers most basic components that can re-appear in several different feeds but also pave the road to several integrations with other specifications. As it can be noted in the summaries of information from real incidents provided by Erka Koivunen[iv] and summarized in the following table, there are some common indicators for most of the incidents. With the use of IODEF data model it is possible to match these indicators with the relative classes and define a first set of metadata that could be mapped after parsing every single feed independently from the source. This allows also to understand eventual gaps in the description of some important indicators that will be covered in the following chapters. In the table below, based on real incidents, most common components are summarized and matched with the IODEF classes in order to give the reader an example of a possible correlation:

Figure 1 - Possible translation of incident indicators into IODEF classes.

Indicator category	**Indicator instances**	**Possible IODEF class**
Actor who has sent the report	Discoverer, Incident reporting clearing-house or Incident Repository, CERT or CSIRT, ISP and Victim	Contact@role
Incident ID	Internal to the organization that is dealing with the event at the moment.	IncidentID
Associated IDs	Related IDs	AlternativeID
Next-in-line incident handling contacts	Incident Repository, ISP's abuse team, national CERT, customer, upstream ISP, downstream ISP and server owner.	Incident@purpose Contact@role
Recorded incident status	Resolved, site taken offline	History
Date discovered (in UTC)	Timestamps	Time
Resolved (in UTC)	Timestamps	Time
Network information	URL, hostname, domain whois, IP whois, IP netblock, ASN, RIR, Multiresolver tests, IP address (PTR), AS name, peer ASNs, country code	System@category Node@category Contact
Actor from whom the evidence was inherited	Discoverer, Incident reporting clearing-house or Incident Repository, CERT or CSIRT, ISP and Victim	HistoryItem
Evidence secured	screenshot , log , links to other public reports by 3rd parties, quick connectivity test, web server vulnerability analysis, malware identifiers	RecordData Reference
Actions taken	Discover, share, receive, verify,	History@action

	analyze, investigate, acknowledge, send takedown request, issue takedown and archive	

An alert can be triggered or a feed can be acquired in various ways and can be composed of several common basic components. All the data coming from internal tools, external sources or incident reports should be gathered together in the system and matched with the other available feeds. In an ideal system data should remain at a local level and should be correlated with external dictionaries regarding to vulnerabilities and attack patterns or behavior information should be exchanged regarding a particular sample of malware. Incident reports should be exchanged via widely recognized standards and not only via email and plain text. The exchange should be automated but also resulting from a push action from / to another response team. Incident reports should be imported/exported via different formats gathering information from all sources available. Once they are received via dedicated and authenticated service, the information should be parsed as for the previous feeds and mapped using a globally recognized taxonomy. This would allow to align the inputs and to have all the information already formatted for a follow up action as the creation of another report to share with other response teams. This process which should be automated should be also subjected to human workflow in order that only when the document is approved, can be sent via the system.

1.3 Outline of the Thesis

The present document aims to investigate the feasibility of this approach and give the reader a full overview on what kind of solution could be implemented within the incident exchange. The goal is to give an insight on the three level of the exchange: the content, the metadata related to the content and the organizational variables that should be addressed in a comprehensive solution.

Chapter 2 - State of the Art: First, it will be proposed a brief overview of the different types of frameworks and exchange formats that can be used to represent or share data about a security event as an incident document or a basic component of the event itself such for example vulnerabilities, malware or other indicators. Various solutions are available to exchange the different components of a incident but none of them addresses the problem in a comprehensive way making use of all available specifications and data.

Chapter 2 - Analysis of a Corpus of Available Data: in order to understand which is the data that

operatively denote an incident information exchange, a mapping of available external data feeds will be proposed. Due to the confidentiality of the information contained in incident exchanges between response teams it was not possible to analyze a corpus of incident notifications regarding single incidents. Therefore it was preferred to focus this initial analysis on widely available information as those that will be here presented.The initial map of the accessible information listed 54 different data providers divided in 8 categories. The categories were malware, honeypots, spam, DNS, SSH, IPv6 and mixed for those sources which provide information that can be reconnected to various categories. For the purpose of this research 23 sources were used to define the corpus of data in this initial feasibility study. The idea was to use the 89 publicly available single data feeds that these 23 data sources provide in order to cover the different possibilities of deployment. The goal was to cover such a diversity of feeds that could allow response teams to extend the data feeds based on their needs. Analyzing the single feeds produced by every examined data source allowed to underline some patterns that were helpful in the definition of the parsing schema and in the definition of an initial ontology conceptualization of incident indicators.

Chapter 4 - Incident Ontology Conceptualization: starting from the 118 values extrapolated from the corpus it was possible to identify five main clusters of incident data in the ontology conceptualization which are incident metadata, address, system, attack and malware. While for incident metadata, address and system a clear definition of the indicators was naturally surfacing, for malware and attacks the definition was not so clear. In filling this gap in the ontology conceptualization three categories from the literature of cyber attack taxonomies were used: attack vectors, operational impact and malware characterization. The goal was to provide a high level definition of the whole incident taxonomy and also propose a starting point for future research especially regarding attack (vectors, impacts and vulnerabilities) and malware characterization.

Chapter 5 - Incident Ontology Implementation using IODEF: starting from the ontology conceptualization presented, it was analyzed the feasibility of an ontology implementation via a widely known exchange format as IODEF. This allowed to note some gaps and possible evolutions of this format that emerged while matching the proposed indicators with the current IODEF data model. These gaps and evolutions lead to some suggestions presented along with implementation indications. These suggestions have been already submitted to the working group in charge of the development of this standard, are now under discussion and if succeed,

will be part of the next version of IODEF.

Chapter 6 - Initial requirements for an application scenario: here the data analyzed was matched with information not incident related in order to cover the full spectrum of the potential requirements of the proposed system. As it was underlined at the beginning the goal of this research is to propose a comprehensive answer to everyday incident exchange needs so it has to cover all the possible data prerequisites. These information cover all organizational and service level aspects of the intermediates involved and also possible data-feed translations and supported standards and formats.

Chapter 7 - The Conficker Example: in order to give the reader an overview of applicability of the proposed model, a brief scenario of use with the data from a widely recognized incident like the Conficker worm was then developed. This malware still represents a current threat in the cyber security scenario and led to the first example of cyber security joint efforts and widely coordinated information exchange.The aim of this scenario was to use available information regarding this worm and frame them in the proposed system showing how this could have been of help in facing this threat.

1.4 Main Results and Future Research

The initial goal of this thesis was to provide an application for sharing information regarding incident for response teams,during this research it became clear that to do it it was necessary to define an overall framework and start to work on a higher level in order to provide a comprehensive solution. What a first sight could have been only an application solution, due to the complexity and the variety of actors involved underlined the need of a definition of first a map of the available solutions, then the definition of the fundamentals of a common language and then the application of a standard compliant solutions and the consideration of all variables involved.These steps revealed the necessity of further research in various areas such a attack and malware characterization and the need to update standard and approaches but most of all, the need of a comprehensive approach paving the road for the standardization of the cyber security exchange. What at the beginning was supposed to be targeted only for response teams revealed the necessity of involving also other actors as data providers, anti-virus company, vulnerabilities reporters and reverse engineers in order to cover completely the different parts of an incident information exchange. Due to the youngness of the cyber security discipline, a comprehensive effort as the one here presented has not been made yet. As for every field the formulation of a

common baseline is part of the beginning and due to the increasing importance of this discipline is every day more required, the present proposal aims to fill this gap and pose the initial questions for such development.It does not pretend to give a definitive question but is the start of a research focused on providing operative answers for all the actors involved. Cyber security today is an intersection of hacking, academia, governments and business and the only way to have a complete picture is to align all these multidisciplinary approaches. At the moment this is hampered by the absence of a common baseline and therefore information framework, the present research aims to give the reader a full overview of the proposed solution and deliver the following outcomes:

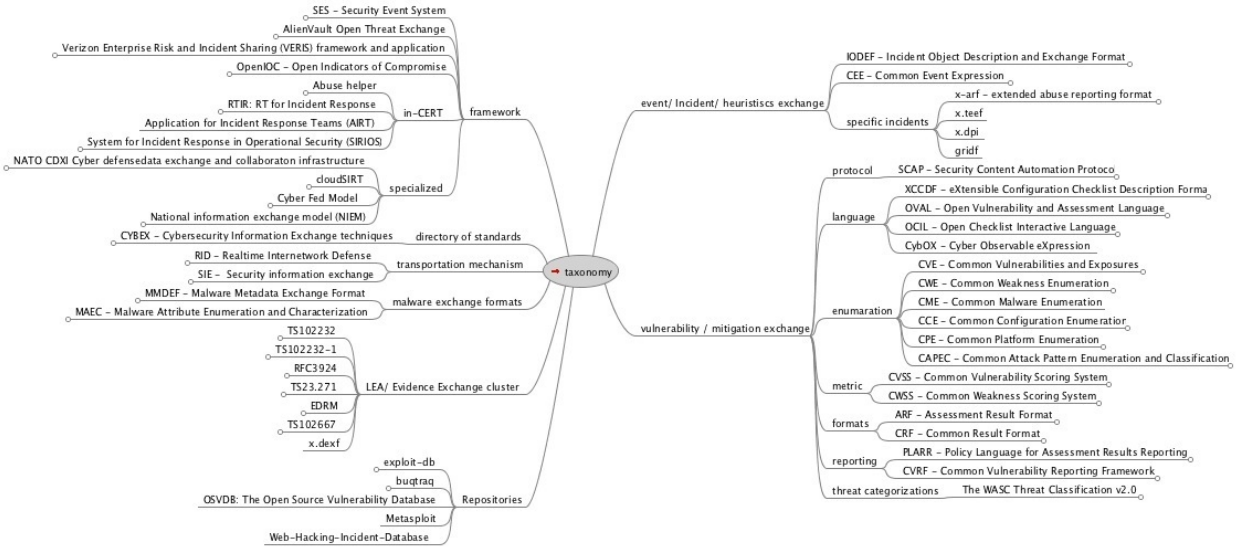
- provide an initial ontology for cyber security incident indicators,
- define the initial questions for the development of a taxonomy for attack vectors, attack impacts and malware categorization that could be widely recognized by the different actors of the security community as response teams, antivirus vendors and security researchers,
- propose the evolution of the IODEF data model in order to improve the adoption of this exchange format and the implementation of the current proposal,
- provide a preliminary study for the future software development of the proposed framework;
- pave the road for an initial standardization of the information exchange for cyber security incidents.

Cyber security information exchange is becoming every day more important, this is due to the increasing resilience of various aspects of everyday life on Information communication technologies (ICT) and also to the crescent maturity of attackers and threats. In order to give an answer to these constantly increasing threats there is the need of a common language and global framework to foster interoperability allowing correlation and collaboration. This approach tackles various paradigmatic issues, requires the consideration of a wide number of different actors, solutions and has a long term horizon. The aim of this thesis is to provide an overview on the first findings of this approach and pave the road for an upcoming research that will cover all the areas that will be presented in the following pages.

2 State of the Art

The aim of this chapter is to give a brief overview of the different types of frameworks and exchange formats that can be used to represent or share data about a security event as an incident document or a basic component of the event itself such for example vulnerabilities, malware or a other indicators. The complete taxonomy includes those efforts that are already operative and those standards that are still under development at the moment. The main idea is to show all the variety of approaches and standards that a response team could use in exchanging an incident. They will be grouped in categories which expand and enrich two prior efforts to categorize specifications made by Dörges[v] and Rutkowski[vi] in 2009 and the Making Security Measurable (MSM)[vii] classification of standardization activities and initiatives.

Figure 2 - Taxonomy of incident solutions and specifications.



As it can be noted the present taxonomy is particularly variegated and an ad-hoc analysis of each specification is not part of the document due to the richness of frameworks, standards and formats available. In order to give a short overview, only the main categories of the taxonomy and the more consistent examples regarding the aim of this thesis are listed below:

- Frameworks: tools that aim to integrate several sources and correlate events in various ways or provide an application to exchange incident documents or parts of it or incident indicators (SES[viii] , VERIS[ix] , Abuse Helper[x] , RTIR[xi], OpenIOC[xii]);
- Event / Incident / Heuristics: specifications to exchange incident related info. Several formats have been defined and are still under definition to provide details about a

security event or generic computer events (IODEF, IDMEF[xiii] , x-arf[xiv]) ;

- Malware: malware can represent one of the basic components of a security event. Available solutions enable to share information about specific threats and theirs behaviors (MAEC[xv] and MMDEF[xvi]) ;
- Vulnerability and mitigation: there are several specifications to share vulnerabilities, weaknesses, system configuration issues and attacks that can be correlated to a security event (CAPEC[xvii], CVE[xviii], CPE[xix], CCE[xx], CVSS[xxi], CWSS[xxii]) ;
- LEA / Evidences: data models for law enforcement and correlation of the incident with evidences and data representations can be part of a security exchange event. (ETSI TS102232[xxiii], EDRM[xxiv]) .

The following schema tries to represent the interaction between internal tools and data from external sources with available frameworks and specifications.

Figure 3 - Interaction between frameworks, internal and external feeds and specifications.

As it can be noted there is a huge gap between the available data feeds and their interaction with system frameworks or use of developed specifications and exchange formats. This is probably due to the ad-hoc development which did not consider all the panorama of available solutions and the definition of specifications in environment that usually are not dealing with operational

issues. Giving an initial answer to this gap is the aim of the present document which starts to tackle the problem analyzing publicly available data from external sources, propose a solution adopting a wide recognized standard and suggest a baseline for a comprehensive framework.

3 Analysis of a Corpus of Available Data

The first step in defining a system for exchanging security incidents is to define the objects of this exchange. In 1998 Howard and Longstaff in "A Common Language for Computer Security Incidents" noted that "there are some other, more general, terms that are required in order to fully describe an incident"[xxv] such as site, date, incident number and corrective action. Thirteen years later in the European Network and information Security Agency (ENISA) report "Proactive Detection of Network Security Incidents" [xxvi] it is underlined: "Basic information the data source should deliver are Autonomous System numbers, IP addresses, domain names, timestamps of incident and of course the category of incident or a set of tags/labels describing it" . Although this data do not fully cover the current threat scenario where for example a vulnerability or a piece of malware could be part of the exchange, it is a useful as a starting point to understand which basic information is more likely to be collected, correlated and therefore shared in a comprehensive solution. Data referring to security events is gathered in two ways:

- Internal tools - tools in direct charge of a response team;
- External sources - data originated by external sources.

3.1 Internal Tools

Every response team has, at different levels, its own internal monitoring system. These systems depend on the tools and technologies adopted, within the networks they supervise. ENISA in the report "Proactive Detection of Network Security Incidents" covers all solutions available. It suggests that the following tools should be part of a response team environment at different levels of its maturity:

Figure 4 - Internal tools.

Standard tools/mechanisms	**Advanced tools/mechanisms**	**Upcoming tools/mechanisms**
Firewall	Darknets	Sandboxing
Anti-virus	Server honeypots	Client honeypots
IDS -IPS	Spamtraps	Passive DNS
Netflow	Networks of sensors	
Application logs		

All these tools provide not only the most common information listed before but also due the

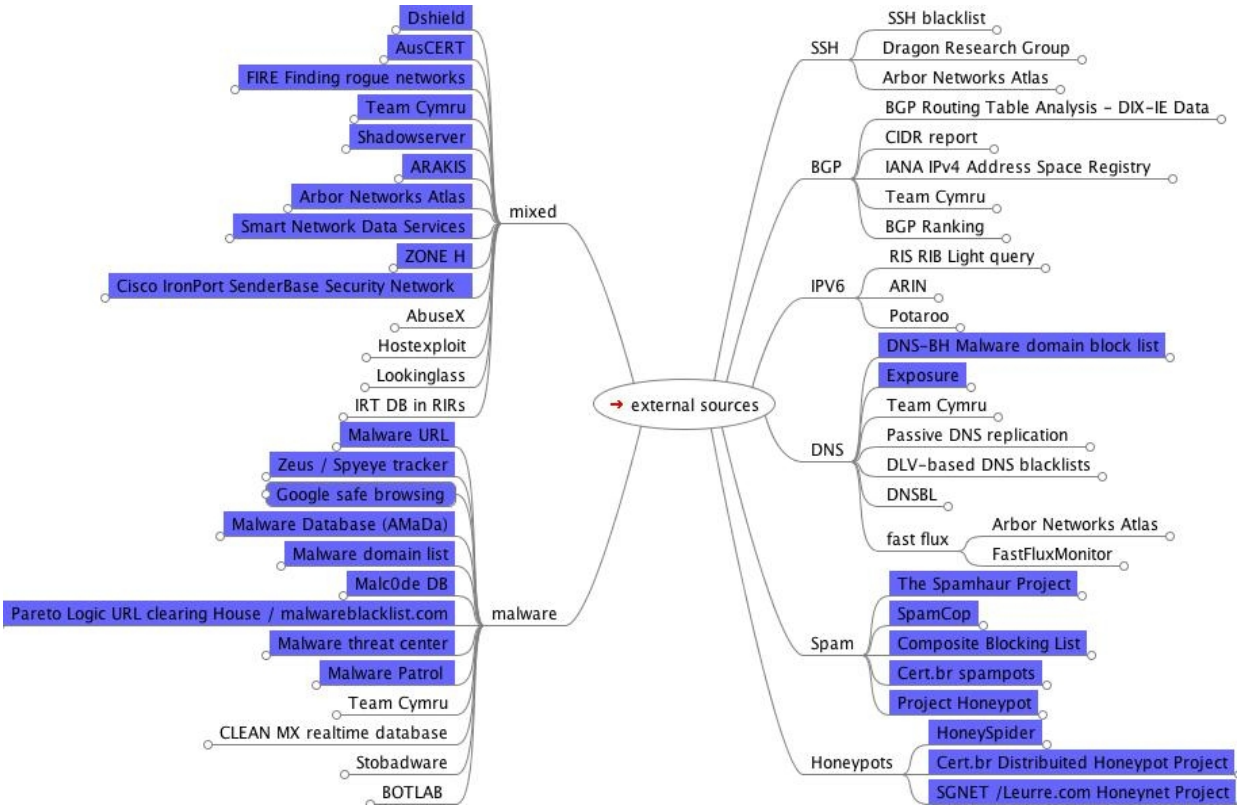
threats they monitor, further additional data that can be correlated with other feeds. Because of their nature, the output formats are mostly proprietary and it is important to underline that they mostly do not make use of particular available exchange formats or specifications. Due to the extensive amount of different solutions and standards, an in-depth analysis of the outputs of these tools will be part of future research.

3.2 External Sources

In a security information exchange data can arrive from another response team in form of electronic incident notification, as a notification from an another external user or as an alert from a data feed provider. In order to investigate which kind of data is available to response teams in the following pages data incoming from external data providers will be analyzed. Due to the confidentiality of the information contained in incident exchange between response teams, it was not possible to analyze a corpus of incident notifications regarding single incidents. Therefore it was preferred to focus this initial analysis on widely available information as those that will be here presented. There are several different feeds publicly available or available under subscription that are an important resource for response teams. Some of the following sources are enterprises while some of them are voluntary based. If from one side this diversity can pose a problem due to the confidence of the data, on the other side it can be seen as a resource to correlate different responses regarding the same basic component. The main differences are the data models and the formats of the output which can range from text (TXT), comma-separated values (CSV), tab-separated values (TSV), Extensible Markup Language (XML) and HyperText Markup Language (HTML) files to SMS and application or system configuration files.

The following table maps all the different kinds of external data feeds. It mostly overlaps with ENISA external data feeds report[xxvii], already cited, but in addition it groups them by categories introducing also data feeds related to SSH, BGP and IPv6 which were not covered in that research.

Figure 5 - External Data Feeds.



The initial map of the accessible information lists 54 different source feeds subdivided in 8 categories. The categories were malware, honeypots, spam, DNS, SSH, IPv6 and mixed for those sources which provide information that can be reconnected to various categories. For the purpose of this research only 23 sources were used to define the corpus of data in this initial feasibility study. The idea was to use 89 publicly available data feeds that these 23 data sources provide in order to cover the different possibilities of deployment. The goal was to analyze such a diversity of feeds that could allow CERTs to extend the data feeds based on their needs. They could add private or internal data feeds or eliminate those they do not consider feasible to their tasks. In doing so this initial implementation aimed to analyze this corpus in order to cover all the possible values and characteristics that they could encounter.

The complete list of the single data feeds and related information is provided in appendix I. Due to specialization of the single feeds they were further grouped in 5 categories:

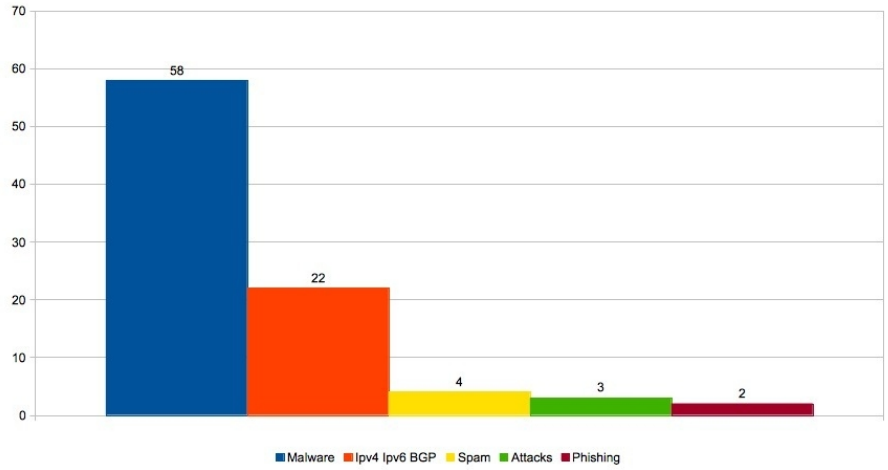
- IPv4, IPv6 and BGP - data regarding allocation and anomalies of the Internet Protocol (IP) v4 and v6 and of the Border Gateway Protocol (BGP) ;
- Spam - notifications regarding addresses which send " mass unsolicited mailings"[xxviii];

- Attack - IP addresses that have been seen as part of SSH, VNC and HTTP attacks ;
- Phishing - notification of URLs involved in "attempting to acquire sensitive data, such as bank account numbers, through a fraudulent solicitation in email or on a Web site, in which the perpetrator masquerades as a legitimate business or reputable person." [xxix] ;
- Malware - this category enumerates website hosting malicious software. Domain and IP addresses hosting botnet, viruses, other related types of software and command & control servers addresses.

3.3 Data Feeds Overview

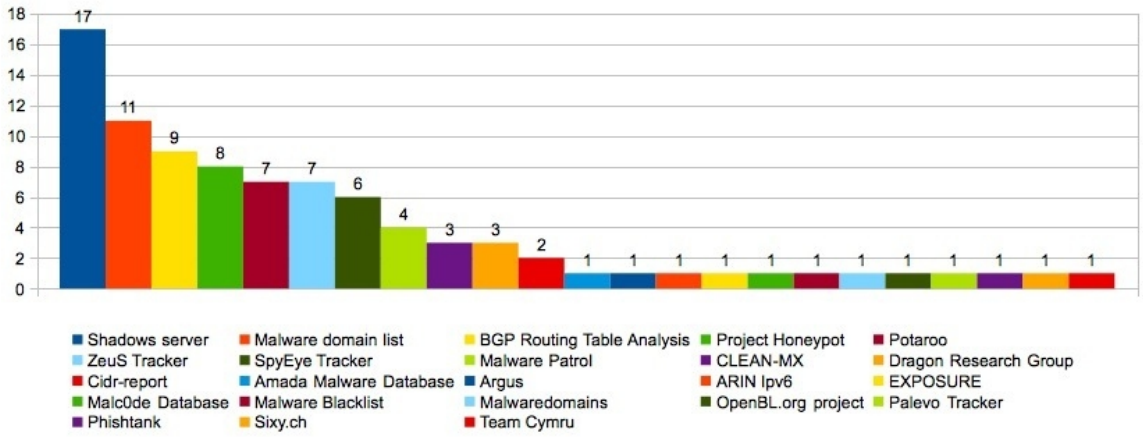
Analyzing the single feeds produced by every examined data source it is possible to underline some patterns that can be helpful in the definition of the parsing schema and in the definition of the following steps like the characterization and correlation. It is important to underline that the present release is a first implementation and sets the base for future developments so it has to cover all different possibilities a response team can encounter in aggregating data publicly available. As it is presented in the table below the majority of the data belongs to malware data feeds wherever they are more related to specific botnets or advertising malware websites. Another important cluster regards IPv4, IPv6 and BGP route advertisement that can be subjected of hijackings and misconfigurations and only a small amount of informations regards attack, spam and phishing. The differences between the clusters are due to the nature of the feeds. For the goal of the present research only feeds publicly available and without submission fee were analyzed. Spam, phishing and attack data feeds are usually subjected to a subscription fee or the owner has to register his/her address space in order to have access to the feed

Figure 6 - Data feeds overview: categories.



The different typologies of feeds influence also the allocations of single data feeds as it can be noticed in the following table where the different numbers of single data feeds provided by a unique data provider are shown. The variety and mixture of approaches while sharing the same data is challenging and also underlines a need of alignment of the different source-specific metadata tags in order to enable the correlation of different instances regarding the same incident. This is due to the fact that there are several efforts overlapping both in type of networks and threats detected but due to the absence of a common language this redundancy is not easily detectable. Moreover in most cases the level of detail is different and information that can be correlated and enrich the view on a particular threat are difficult to identify again due the absence of a common language.

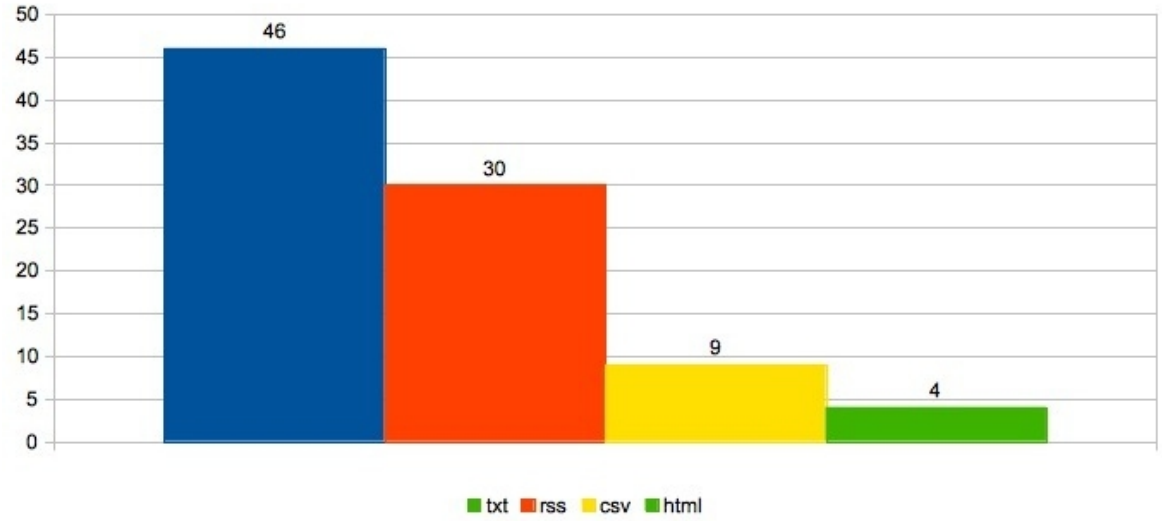
Figure 7 - Data feeds overview: single feeds per data provider.



Another important feature is the output for the feeds. Usually the feeds are available in different formats but for the purpose of this research, the preferred formats were the ones with less

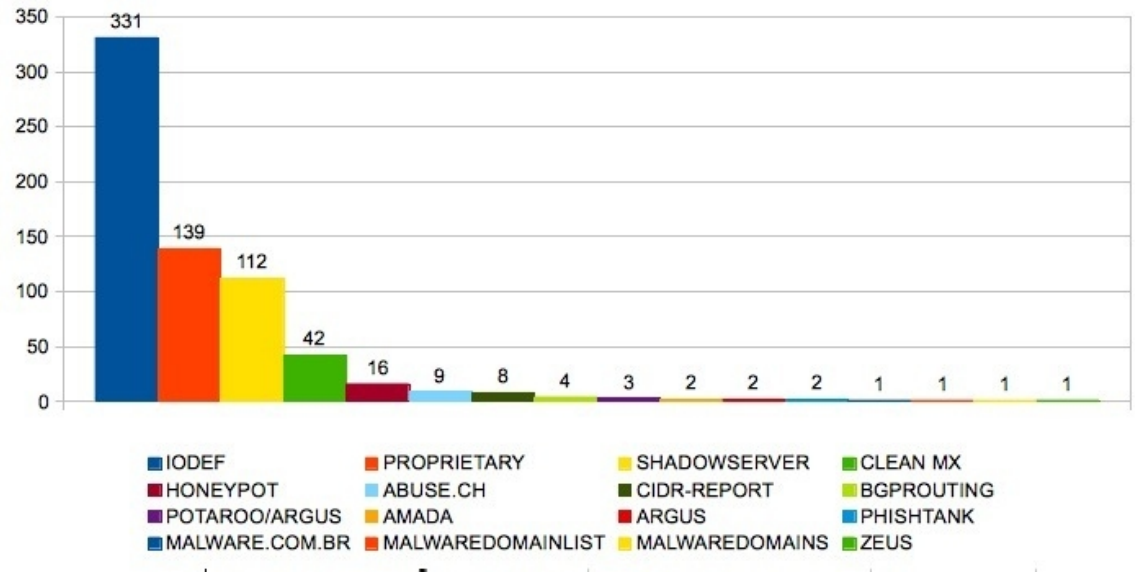
overhead details and easier capability to be parsed. In the following table the distribution of the single data feeds is represented and as it can be noticed 86 % of the feeds can be directly parsed from a plain text file. Due to the richness of contents that can be found on the Internet and the large deployment of these formats, it is important to consider all the different options in order to develop a system flexible enough to cover all needs of aggregation.

Figure 8 - Data feeds overview: formats.



Typology, data sources and formats play an important role in a definition of such a framework. As expected, even for data regarding same category types, several differences regarding the labeling of the metadata were found. Contents aggregated by such a large number of single data feeds provide a perfect overview on the need of a common vocabulary even of the main basic components. In order to understand how data from these various feeds could be grouped every single value was analyzed. When possible it was grouped with similar values under the same label. Each feed has an average of 8,5 fields and the total single values were 674. Starting from this wide base it has been possible to reconnect all the feeds to a total of 118 metadata which try to cover all the different characteristics of the data analyzed. In appendix II - List of the metadata values - the complete registry of all values it is provided. The list is organized in alphabetical order and for each metadata value provides the code of the value, the description and the source that can be IODEF, proprietary or reporting the source of the native metadata of the feed where it was not possible to group it with IODEF or a proprietary metadata. In the following table the distributions of the occurrences of every kind of used values are reported:

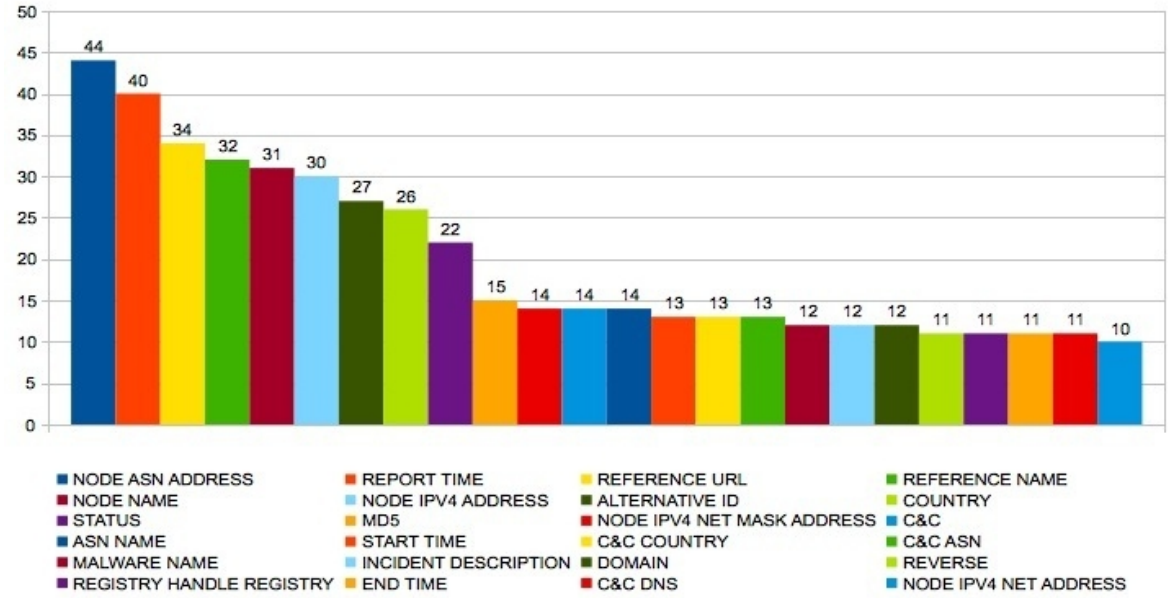
Figure 9 - Data feed overview: single values occurrences.



As it can be noticed the variety of different contents and therefore metadata provided influenced the definition of the single values. While some important metadata values were possible to be reconnected to IODEF data model, others were specifically related to the nature of the feeds and nor IODEF or proprietary values could be matched. It is worth to notice that IODEF do not provide specific metadata for malware, DNS or specific geographic fields in general or command and controls characteristics in detail. This lack of details regarding most of malware related fields will probably foster the interoperability with other standards currently in development such as Malware Attribute Enumeration and Characterization (MAEC). Where metadata related to basic information as domain data or attack types were not found in the IODEF, they were then grouped under metadata named as proprietary. As it can be noticed half of the occurrences were reconnected to the IODEF data model, while the other half was divided between proprietary metadata labels and source specific values. The source specific values will be subjected to further investigation in order to understand if they can be reallocated to already existing fields or future development of other specifications. In the following table are presented the first 24 metadata values or fields labels that are mostly occurring in the analyzed data set. They can be sub grouped in Node and C&C (IPv4,BGP and DNS), reference and ID, timestamps, malware and description characterizations. It is important to underline that these values represent the main indicators in a record but they are not always present in every feed. Dealing with feeds from different data source providers and different typology implies an approach that enables to correlate item with different values but sharing the same common

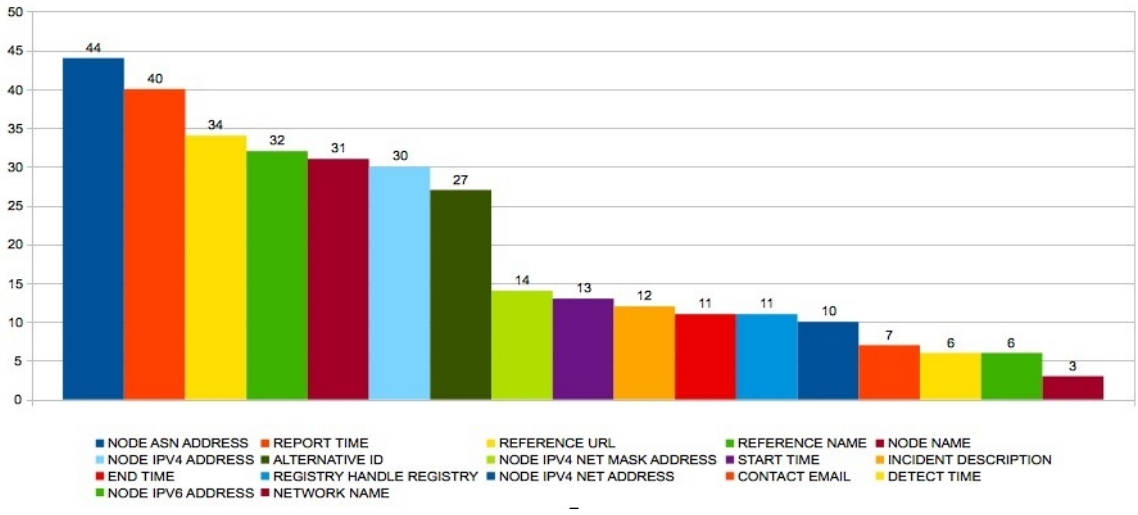
characteristics.

Figure 10 - Data feeds overview: typology of metadata.



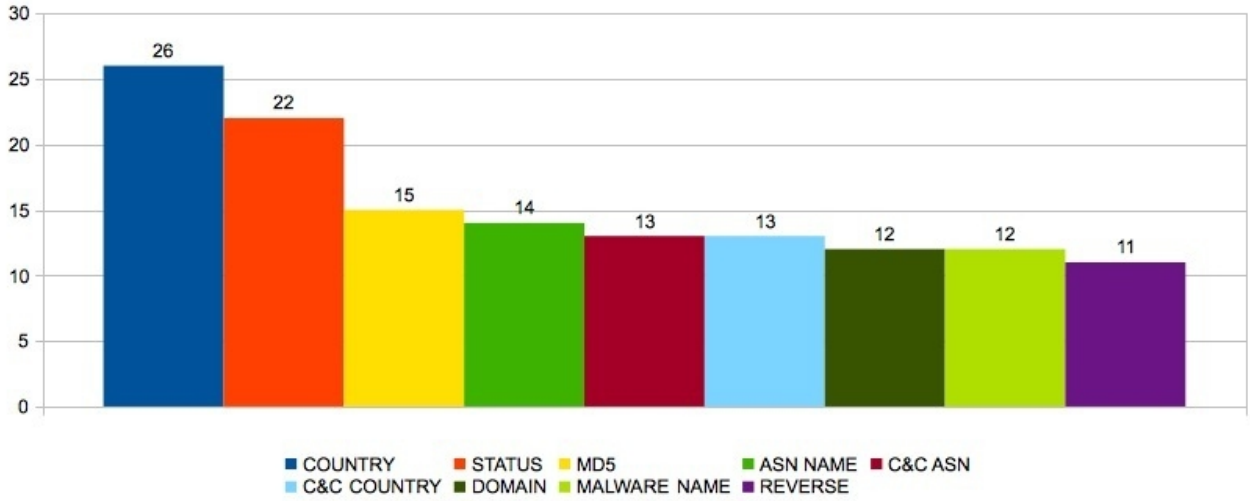
While trying to resemble connecting patterns between different flows of information it is important to start normalizing using existing standards as it was mentioned before, this is why the IODEF data model was used as baseline. In the next table are underlined which one of the above mentioned main fields can be reconnected to this data set. As it can be noticed they are half of the most used values presented in the previous tables and they cover the main characteristics of an incident evidence such as node, id, time and description fields.

Figure 11 - Data feeds overview: main fields.



Where it was impossible to use fields derived from the IODEF dataset and various values could be grouped under the same metadata of a more generic label, this was done using a proprietary field. The idea behind this categorization is to group values of the same categories under a common label that can be renamed in the future due to extension of IODEF or other specifications that are under development at the time of writing. In the table below are summarized the occurrences of the proprietary values:

Figure 12 - Data feeds overview: occurrences of proprietary values.



As it can be noticed these are high level details regarding a network address or malware instance. The idea is to convert them in a more appropriate standard compliant field as soon as the new specifications or IODEF extensions will be deployed. Further implementation and variations will be possible within the development of the system or the usage of other data models. It is in the aim of the research to minimize the use of source-native fields and to understand the implementation of standard compliant fields.

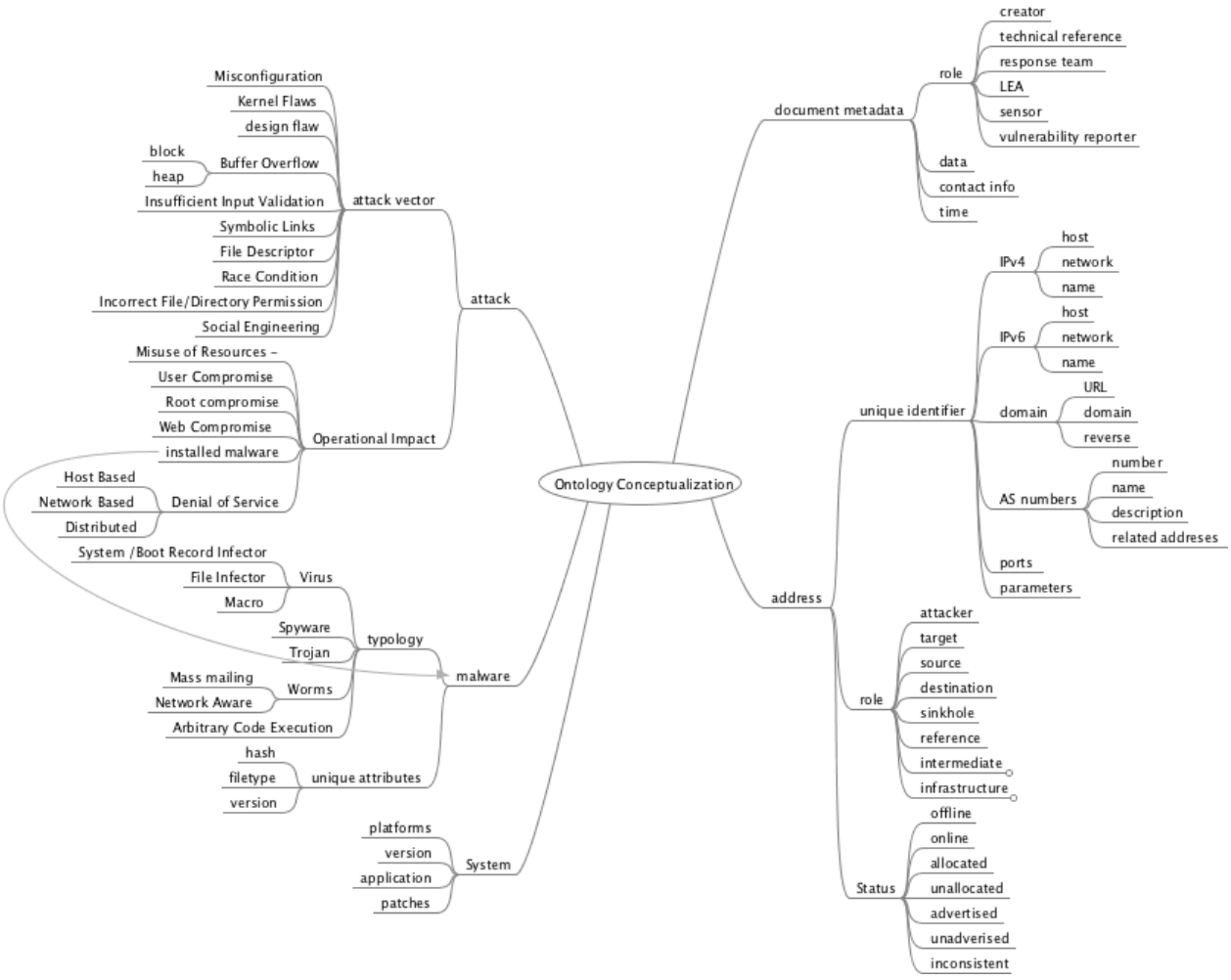
4 Incident Ontology Conceptualization

In the previous chapter all the different publicly available feeds were presented, as it was possible to notice, most of the feeds had several commonalities like IP address, timestamps, URL or domain, ASN numbers and other information mainly regarding malware or attack data. It is clear that when it comes to the description of a threat there are some common indicators that are shared among all the security community. The gap resides in the fact that there is not common vocabulary or practices how this data is defined and labeled. At the moment there is no globally recognized language or standard widely used for incident exchange, as in a sort of babel everybody is talking about the same elements but due to a lack of standardization incident information sharing is hampered. It is possible to parse every single feed in a manual or semi automatic process but there is no global acquired high level taxonomy that could leverage the practice and provide a base for mutual exchange. The idea of this chapter is to propose an "ontology" where to map all the atomic elements or "indicators" as Daley, Millar, & Osorno define these single elements in the "Operationalizing the Coordinated Incident Handling Model" paper written in 2011[xxx]. The creation of common indicators is important also in other critical fields, as Andrew Handale in his speech at the Securities Industry and Financial Markets Association symposium reported: "The economic costs of this linguistic diversity [in the financial markets] were brutally exposed by the financial crisis "[xxxi].The paragon between financial exchange and security exchange can seem reckless but considering the possible threats that are arising, pave the road to a common language can be of help in facing the next generation of security incidents. The need of a common vocabulary is not only necessary for sharing security information but also for post mortem analysis. As a title of example in 2012 there is no global statistics regarding computer incidents across the world, all reports display just a minimal part of the current threats that everyday security response teams deal with. While incident sharing is raising as an important item as recently exposed by the US Cyber Intelligence Sharing and Protection Act [xxxii], the need of a clear vocabulary is emerging only in some areas. The purpose of this chapter is to extrapolate an Ontology Conceptualization [xxxiii] from the metadata emerged in the analysis of the external data feeds, underlining eventual gaps in the current scenario and proposing a starting point for future research especially regarding attack (vectors, impact and vulnerability) and malware characterization.

4.1 A Proposal for an Ontology Conceptualization of an Incident

In "Appendix II - list of metadata" values are enumerated all the different values that can be acquired from external data feeds, a first attempt to categorize them was made and had lead to 118 values. The goal now is to aggregate the data in a first ontology conceptualization that will pose the base of the ontology implementation via IODEF. Starting from the values extrapolated from the corpus, an initial aggregation of the values was made then a further abstraction in order to identify the most basic elements of the exchange was performed and here is provided.

Figure 13 - Proposed incident ontology conceptualization.



For each element and attribute is proposed a definition. In order to provide an authoritative record, where possible the definition submitted is taken from academic literature or Request for Comments (RFC) literature[xxxiv]. The goal is to provide a standardized vocabulary for incident exchange in order to give a quick reference not only for developers or incident handlers but also

for other interested parties. Once the fundamental components of the exchange are defined it is then possible to enrich the content with all the detailed information that can be thus provided. The identified main clusters of an incident are five: incident metadata, address, system, attack and malware.

4.1.1 Incident Metadata

The incident metadata is composed by the following elements: role, data, contact info and time.

- Role - it indicates the role of the intermediaries within the information exchange:
    - creator - the creator of the feed or the document;
    - technical reference - the technical reference for the host or network;
    - response team - the response team to whom is sent / or which received the document;
    - LEA - the Law Enforcement Agency involved in the incident;
    - sensor - the entity that generates the internal / external feed;
    - vulnerability reporter - the entity who reports the vulnerability.
- Data - the data attached to the document for example logs, audit, samples.
- Contact info - all the contact information (registry, email and postal address, phone numbers, Pretty Good Privacy (PGP) [xxxv]keys)
- Time - date time information (start, end, reported, detection, timezone)

This cluster describes the main generic characteristics of the information exchange. These indicators vary from relationships with the data, the generic contact information related to the provider of the data or the data itself, the temporal definition of all indicators to the description of the data included in the information exchange. Moreover this cluster is going to contain all information regarding particular restrictions or exceptions that will be required the more incident response involves privacy issues and transnational incidents.

4.1.2 Address

The address class is organized in:

- Unique identifiers;
  - IPv4 /IPv6:
    - host - host or computer of an end user is intended to refer to a computing device that connects to the Internet[xxxvi] ;
    - network - the group of host that are interconnected under a certain network;
    - name - the generic name of the host or network.
- Domain Name System:
  - URL - Uniform Resource Locator as defined in RFC 3986[xxxvii] as for example: http://www.math.uio.no/faq/compression-faq/part1.html ;
  - domain name - the fully qualified name of a specific domain;
  - reverse - the reverse domain or pointer (PTR) record.
- AS numbers - BGP protocol:
  - number - Autonomous System Number;
  - name - name of the legal entity in charge of the AS;
  - description - Autonomous System Number description;
  - related addresses - network addresses associated to the Autonomous System Number
- Ports - logical entities for Internet communication[xxxviii] and related IP protocols;
- Parameters - specific values required by the protocol;
- Role:
  - attacker - threat agent [xxxix] ;
  - target - a computer or network logical entity (account, process, or data) or physical entity (component, computer, network or internetwork)[xl] where an action of compromising is focus to;
  - source - a system originating the traffic;
  - destination - the destination to which the traffic is sent to;

- sinkhole - a system where the traffic is redirect;
- reference - a system where metadata regarding an element is contained;
- intermediate - a system which is involved in the path;
- proxy - a system that acts in behalf of another system;
- infrastructure - a system which is needed for the existence of another system.

- Status:
  - offline - not connected to the Internet/s;
  - online - connected to the Internet/s and reachable at a certain moment;
  - allocated - delegated entirely to specific RIR[xli] ;
  - unallocated - not yet allocated or reserved. [17];
  - advertised - advertised in BGP[xlii];
  - unadvertised - not advertised in the routing system [18] ;
  - inconsistent - inconsistent route or address.

This is the main and most detailed cluster and indicates the characteristics of the location originating, involving or targeting by the reported event. The declinations of the indicators regarding an address are variegated since they can range by category (IPv6, IPv4, DNS, BGP), by cardinality as host or networks, typology (source, target, attacker, proxy, sinkhole, C&C) and status (offline, online, allocated, unallocated, advertised, unadvertised). Other generic informations are not depicted here but they encounter geographical position of the hardware hosting the host/network and other address specific variables. This cluster encompasses all Internet unique identifiers such as domain name, IP, AS numbers, ports and parameters and their characteristics and role in the incident which is subjected of the information sharing.

4.1.3 System

The system cluster is organized in:

- Platforms - operating system;
- Version - version of operating system / software;

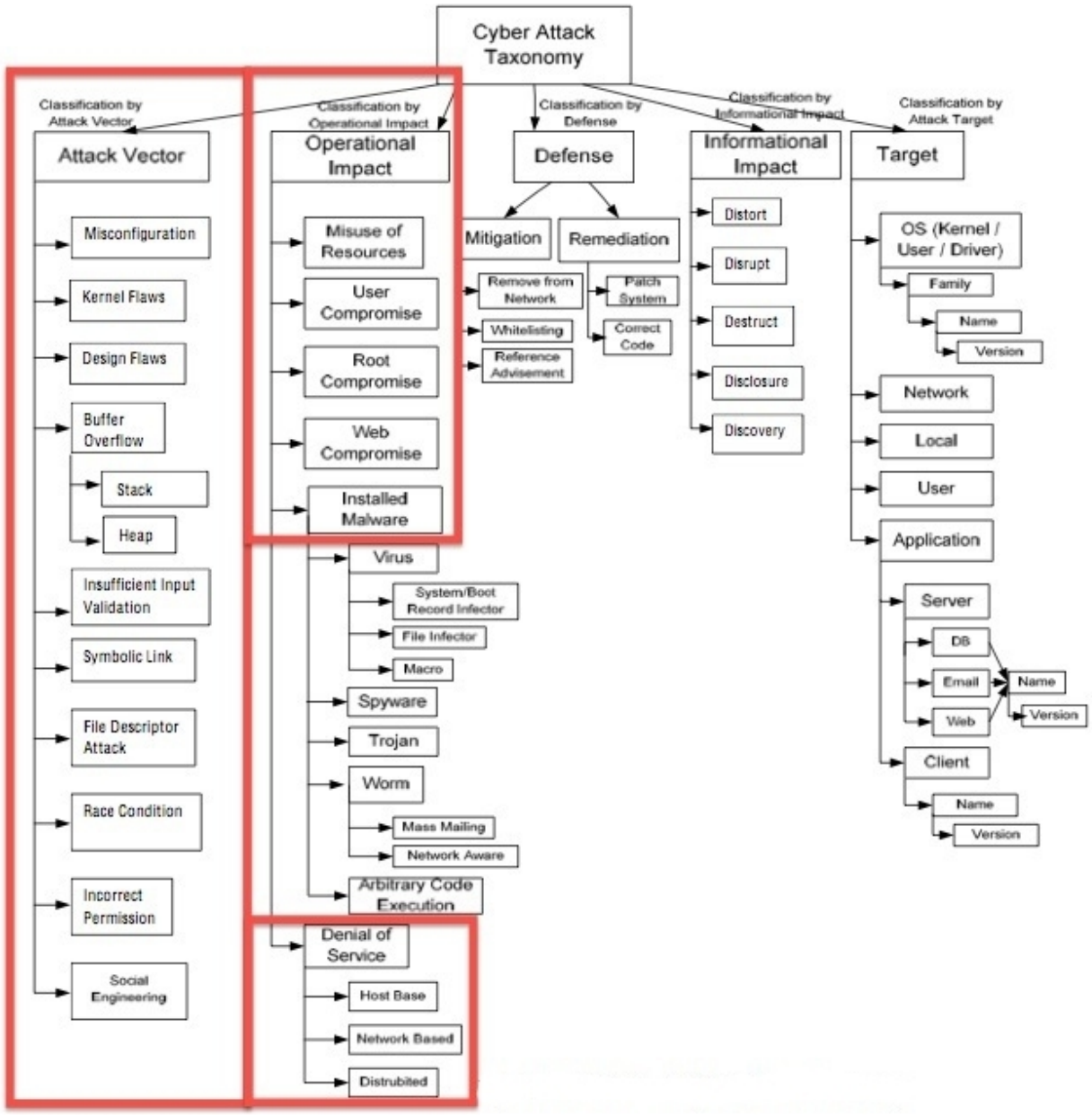
- Application - specific application;
- Patches - patches available.

As it can be noticed in the data feeds presented, information about the systems involved are usually exchanged on a high level of description. Nevertheless is important to assume that in internal data feeds or incident document exchange these indicators not only provide information about the operative systems involved but also specific indicators relative to platforms, versions, application and relative patches.

4.1.4 Attack

While for addresses and applications a clear definition of the indicators was naturally surfacing, for malware and attacks the definition was not so clear. Several attempts[xliii] [xliv] [xlv] have been made over the years regarding these two clusters but none of them have been widely adopted. In defining this ontology conceptualization two categories from the "cyber attack taxonomy AVOIDIT"[xlvi] will be borrowed: attack vectors and operational impact were the sub cluster installed malware will be further developed as a cluster freestanding. The goal is to provide a high level conceptualization of the whole incident taxonomy. As for the International Union of Pure and Applied Chemistry (IUPAC) nomenclature, the aim of this conceptualization is to define in an unambiguous way all the elements of an incident exchange, starting from the highest level of description and deepen via ontology implementation using all available detailed indicators currently available. As anticipated, the attack cluster is divided in attack vector and operational impact and in this initial proposal the AVOIDIT proposed taxonomy will be used:

Figure 14 - AVOIDIT cyber attack taxonomy: attacks vectors and operational impact.



The attack vector characterization is proposed here as a normative example in order to underline how research in this direction should be fostered. AVOIDIT defines an attack vector as “a path by which an attacker can gain access to a host”. Moreover in this definition it includes also vulnerabilities. For the extent of this research this topic will be not expanded in this document since it is clear that it needs ad-hoc research. Finding an agreed way to define an attack vector and therefore a vulnerability can be not only a benefit for the incident response community but also for the entire security community. This is due to the fact that security researchers find and report vulnerabilities to vendors/response teams and vendors patch vulnerabilities for end users

while attackers exploit vulnerabilities to gain access. Another important characterization is the one regarding operational impact These indicators describe the interdependencies between addresses and malware and help in further categorize the incident in action. In doing this they help in framing not only the the action subjected to the information exchange but they also provide an interpretative key for the possible interconnections and effects involved. There are two more main points to underline regarding the attack cluster. Considering the volume of these events, not having globally accepted categories hampers the alignment of incident categorization. It also hampers to compare statistics between different response teams and have a global view on the real numbers. Implementing a general characterization could help as a first step in this direction. Considering that every response team has individualized vocabularies but at the same time there are several cyber security enumerations, scorings and formats regards attack patterns, it could be useful to have a high level enumeration and definitions which response teams can use to map their own dictionaries and where to insert their internal characterizations and names as a subset. Moreover a general registry could be seen as a first effort in the direction of a standardized high level characterization for attack and could be then linked both to more organized frameworks and also proprietary implementations. The provided taxonomy taken form the AVOIDIT framework is just an initial proposal. It is clear that in the absence of a globally recognized taxonomy any attempt to use a proposed one it will probably subjected to change. The main goal regarding this research is not to provide a definitive answer to the gaps in the incident response but to underline which are the gaps that should be filled and which improvements can be suggested in order to broader the adoption of this framework and enhance information exchange.

4.1.5 Malware

The Malware cluster involves all the indicators gathered to uniquely identify a specific sample of malware. The unique identifiers are used to define the unique pieces of malware which are part of the information exchange (hash, filetype and version) and the typology helps to identify the type (botnet, rootkit, exploit, worm, trojan, backdoor, adware, key logging, click fraud, virus, etc).These indicators are not immediately present but are emerging as long as the incident analysis is performed.

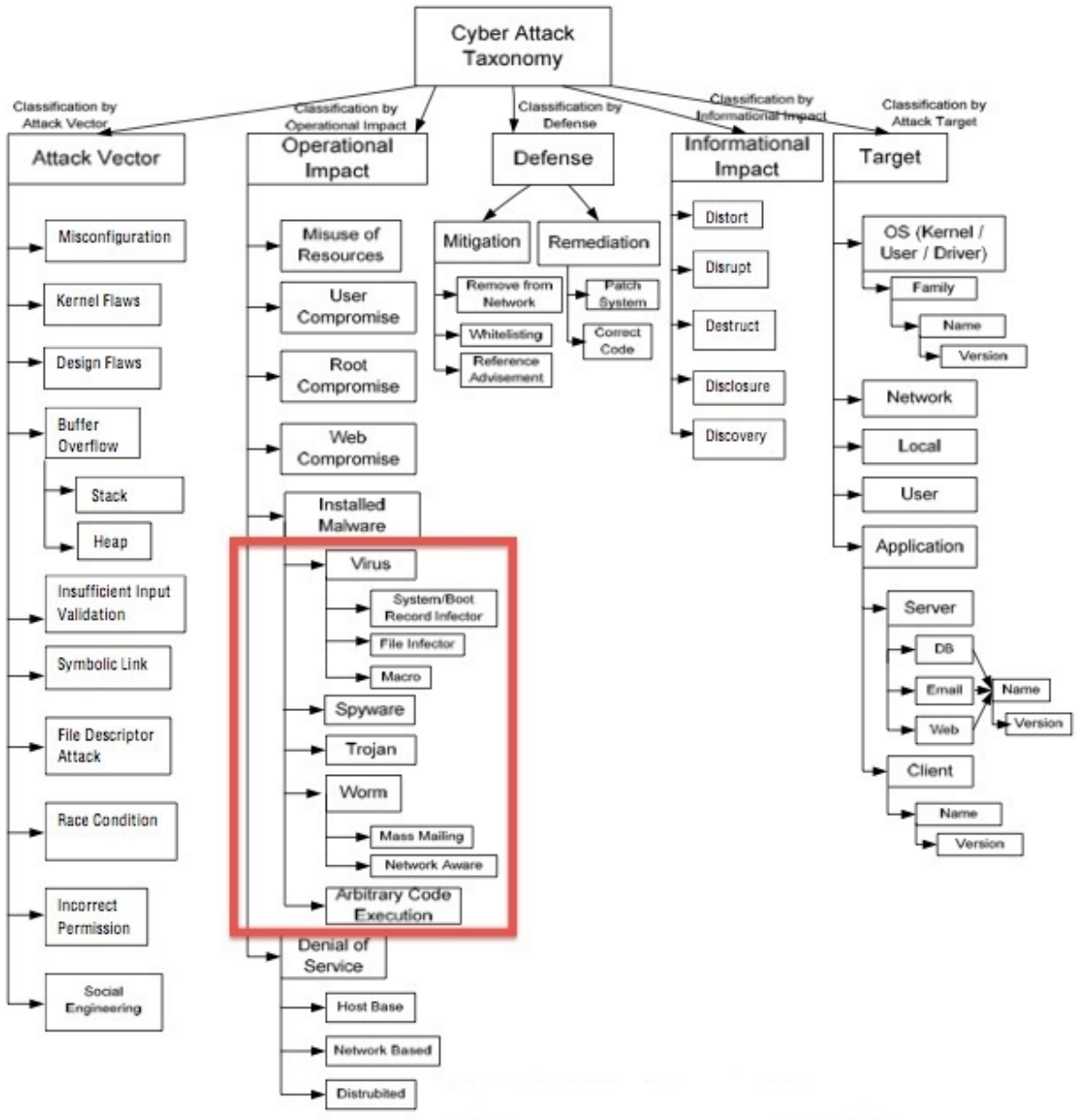
Unique identifiers:

- hash;

- filetype;
- version.

Typology:

Figure 15 - AVOIDIT cyber attack taxonomy: malware characterization.



As for the attack vectors and operational impact this characterization is not exhaustive and a deeper research should be acknowledged in order to define a globally recognized taxonomy. These indicators are directly connected to the attack vector type and help in defining the nature of the incident. The description of a malware sample can be very detailed but at this stage of the

taxonomy only high level indicator should be considered. The proposed schema uses the AVOIDIT taxonomy for malware but as was valid for the attack vectors and impact this is just a initial proposal. and will be probably subjected to changes after future research. As for attack, every single response team or anti-virus company or vendor has its own internal dictionary and so there is no globally recognized high level malware taxonomy. For these reasons a generic characterization should be developed in order to allow interoperability and data exchange. There are some efforts in this direction but or they are not updated or are too focused on detailed description and will be covered in the next chapter. Moreover due to the presence of several malware related data feeds as exposed in the previous chapter , every data provider have its own way to characterize even the same sample. Considering the diversity of actors involved as data providers, anti-virus company, response teams, reverse engineers and end users the definition of a straightforward, updated and globally acquired taxonomy could only be a benefit for all. Another important part of the malware indicators are the artifacts or the traces that malware leaves of its presence in systems (registry keys, files and mutexes[xlvii]). These indicators although were not contained in the analyzed data feeds represent an important part of malware research and will be integrated in the future phases of this project which aims to correlate all possible indicators of infection.

5 Incident Ontology Implementation using IODEF

In the ontology conceptualization the main generic indicators used in the information sharing were depicted. These indicators were high level and can be of help to define the main characteristics of an incident. The aim of an ontology implementation is to analyze the feasibility in adopting one exchange format as IODEF and to underline eventual gaps or evolutions of this format that could emerge during deployment. The goal is to give a practical example of implementation of the model and therefore pave the road to the adoption of this format and also of the presented approach.

In order to categorize the different events, various exchange formats have been defined and are still under definition. They differ one from another mainly by the object they consider. The most comprehensive is IODEF which was developed by Internet Engineering Task Force (IETF) based[xlviii] on Intrusion Detection Message Exchange Format (IDMEF). IODEF is defined in RFC 5070 and represents an Extensible Markup Language (XML) data model to transport information about security events. This data model enables a comprehensive definition of various basic components of an incident representation and embodies an important step in the definition of a common syntax. IODEF has been extended in 2010 by the Document Class for Reporting Phishing defined in RFC 5901[xlix] while other several extensions are under development by the IETF Managed Incident Lightweight Exchange (MILE) Working group. Other important efforts in incident indicators exchange are OpenIOC - Open Indicators of Compromise[l] developed by Mandiant which sets the ground for the recent defined new standard language for cyber observables, Cyber Observable eXpression (CybOX)[li] from MITRE and the Verizon Enterprise Risk and Incident Sharing (VERIS) framework and application.

5.1 IODEF and its Extensions

From the analysis of the single data feeds it was possible to see that no data provider except one[lii] was making use of any particular exchange format. The available specifications or are completely unknown to the data providers or are considered too complicated to use. None of the them is globally used or spread or officially recognized apart IODEF and its extensions which at least are currently in RFC standards track. RFCs in standards tracks are expected to be "characterized by a high degree of technical maturity and by a generally held belief that the specified protocol or service provides significant benefit to the Internet community."[liii] For the

purpose of this thesis the data model of theses two standard track RFCs:

- RFC 5070 - The Incident Object Description Exchange Format - December 2007
- RFC 5901 - Extensions to the IODEF-Document Class for Reporting Phishing - July 2010

and the following drafts in discussion at the moment were analyzed[liv] :

- IODEF extension to support structured cyber security information[lv] (Internet-Draft)
- Expert Review for IODEF Extensions in IANA XML Registry[lvi] (Internet-Draft)
- Guidelines for Defining Extensions to IODEF[lvii] (Informational)

The goal was to back up the ontology conceptualization previously provided with an international recognized specification as IODEF and then define the parsing model and the following initial requirements for a possible software development. The use of this standard should be almost transparent to the response team in charge of the information exchange. The goal is to provide the fundamentals of a system which is compliant to a recognized standard but has the flexibility to interact with different technologies. In analyzing the feasibility of the approach the writer mapped some gaps and potential evolution that could be applied to IODEF. In order to pave the road to the implementation of the present research, the suggestions that are presented in the following paragraphs have been already submitted to the MILE working group and are now under discussion.

5.2 The Incident Ontology and the IODEF Data Model

For each one of the main clusters presented in the ontology conceptualization a possible use of existing IODEF classes or a possible evolution or extension of the current data model will be presented. Moreover it will be also detailed a possible implementation with other cyber security exchange formats and specifications.

5.2.1 Incident Metadata

This is the data referred to the document or the data feed that is gathering the information.

Role: IODEF covers most of the presented information within the contact@role class. There are some values that are not directly mapping with this class but this can be done via cross reference with the system@category for what concerns the sensor value. For what concerns LEA and

vulnerability reporter there are two ways of implementation:

- implement the escape value to extend this attribute;
- foster the adoption of these values within the official data schema.

Data: various data samples can be attached to the notification. This data is fully covered by the RecordData class and related subclasses.

Contact info: considering the information in this class, IODEF covers all the possible information. It must be underlined that due the extensive use of PGP within the security community an ad-hoc element of the class should be considered for future implementation.

Time: the time classes offered by the IODEF cover most of the values encountered but there is one value that occurred in the corpus of data which was not covered which is next validation time. This gap can be fulfilled with escape value.

Available interoperability with other cyber security exchange formats and specifications: There are no particular cyber security information or particular specifications regarding this cluster except for those data concern timestamps. In this regard the following RFC 3339 "Date and Time on the Internet: Timestamps"[lviii] is used as a reference.

5.2.2 Address

Unique identifier: to define an address IODEF already includes all the granularity to define IPv4, IPv6 and ASN numbers in the system and node classes. At the moment only domain data is not fully covered by the current data schema and in order to fulfill this gap two actions are possible:

- to use the definition of the class from RFC 5901 and describe the information with the DomainData, DomainContacts and Nameservers elements;
- to implement a new class regarding the domain data with an update of the IODEF data schema. This action is currently in discussion within the MILE mailing list and if it will succeed will be part of the new data model.

Role: currently IODEF covers the typology of role of the system@category enumeration. The present values are infrastructure, intermediate, sensor, source and target. While attacker and destination could be considered as covered by source and target, they probably don't completely comply with the definition within RFC 5070. Another important point regards sinkhole and

command&controls, these specific roles and possibilities are not clearly described in this RFC. The question was raised within the MILE mailing list and can be addressed in two ways:

- a new guidance of the current data schema using existing values;
- the further implementation of the available values in the system@category.

Status: At the moment IODEF does not handle the status of an address as detailed in the ontology conceptualization. In the RFC 5901 there are two enumerations regarding the Domain status and System Status that can partially fill this gap. Unfortunately they do not reflect the values extrapolated from the taxonomy which are the following:

- offline/ online
- allocated /unallocated
- advertised/ unadvertised
- inconsistent

There are two actions that can be done:

- fostering the implementation of a new guidance of the RFC 5070 system@spoofed (yes, no, unknown) in accordance with RFC 5901 values for system@status (spoofed, fraudulent. innocent-hacked, innocent-hijacked, unknown) and domain@status (reservedDelegation, assignedAndActive, assignedAndInactive, assignedAndOnHold, revoked, transferPending, registryLock, registrarLock) which occurs only in Fraud reporting;
- implement a new set of enumeration which expands RFC 5070.

Available interoperability with other cyber security exchange formats and specifications: there are no specific cyber security specifications, standards or exchange formats regarding how to express the characteristic of an address and its unique identifier except those define in RFC 3986 “Uniform Resource Identifiers (URI): Generic Syntax" .

5.2.3 Attack

The attack part is not covered at the moment in IODEF. Two kinds of attempts can be recognized in this direction. One is the Impact@type values of RFC 5070:

- admin. Administrative privileges were attempted;
- dos. A denial of service was attempted;
- file. An action that impacts the integrity of a file or database was attempted;
- info-leak. An attempt was made to exfiltrate information;
- misconfiguration. An attempt was made to exploit a mis-configuration in a system;
- policy. Activity violating site's policy was attempted;
- recon. Reconnaissance activity was attempted;
- social-engineering. A social engineering attack was attempted;
- user. User privileges were attempted.

and one is the Fraud@type in RFC 5901:

- phishing;
- recruiting;
- malware distribution;
- fraudulent site;
- dnsspoof.

As it can be noted this characterization lacks of structure due to the overlap between attack vectors and impacts and hampers a clear identification of the occurring incident within the IODEF data model without the use of the Fraud extension. Indeed the more detailed description provided by RFC 5901 occurs only in case of fraudulent incidents. Due to this lack of equivalent values between the provided ontology and the IODEF data model, it is currently under discussion in the MILE mailing list whatever

- further guidance document should be provided on how to use the existing IODEF data-model possibilities to characterize an attack;
- the definition of an high level taxonomy should be discussed and implemented.

Available interoperability with other cyber security exchange formats and specifications: helping in the definition of an incident there are three main cyber security structured indicators

related to attack patterns, vulnerabilities and weaknesses that can be used. As proposed in the draft "IODEF-extension to support structured cyber security information" IODEF can be extended using the Common Attack Pattern Enumeration and Classification (CAPEC) for what concerns the typology of the attack. The current taxonomy covers 400 attack patterns subdivided in 68 categories. As it can be deducible by these numbers a high level taxonomy which enables a quick reference both for the handler and the reporter is needed as proposed in this research. Regarding the attack vector, the same draft already foreseen the possibility of the implementation with indicators from Common Vulnerabilities and Exposures (CVE) which is widely recognized as probably the most structured and popular way to reference a vulnerability. Another interesting extension is the implementation of the Common Vulnerability Reporting Framework (CVRF)[lix] which tries to leverage the disclosure of vulnerabilities. CVRF can represent an important help in aligning the disclosure of a vulnerability that a response team need to know and it is worth to mention that Microsoft started to support it in all its monthly security updates[lx] during the writings of this thesis. But it must be also considered that this effort do not cover the whole panorama of vulnerability disclosure. In this concern it must be underlined how vulnerability are mostly disclosed on free form mailing list and are also subjected to several different proprietary disclosure formats. In 2002 an attempt to define a method on the disclosure process was developed via the draft "Responsible Vulnerability Disclosure Process"[lxi] but the proposal have expired. In order to try to cover as much vulnerabilities as possible it must be considered also the implementation of extension to exploits repositories such as Open Source Vulnerability Database[lxii] and The Exploit Database[lxiii]. Other possible indicator that can extend the IODEF data model and enrich the set of information regarding an attack can then be Common Vulnerability Scoring System (CVSS). It can be also extended by the Common Weakness Scoring System (CWSS) which is associated with the Common Weakness Enumeration (CWE)[lxiv] which enables the enumeration of set of weaknesses. In this respect it must be underlined that there are also other scoring systems, such as Microsoft[lxv] for example but they are proprietary. While the identification of widely adopted cyber security extensions within IODEF is important, these proprietary indicators should be also considered in a comprehensive framework.

5.2.4 Malware

Unique identifier and typology: IODEF does not include any specific class to malware samples.

The reference class can be used to refer to a particular sample but there are no related enumeration or specific classes to identify which kind of malware infections occurs in the incident. In the RFC 5901 a first attempt is made both with the fraud@type value "malware distribution" and also with the relative class "included malware" and the subclass data, the "FilesDownloaded" and "WindowsRegistryKeyModified" class. Considering the importance and the volume of these threats in the current scenario there are two ways to fulfill the gap:

- improving the guidance of the current IODEF data model and related extensions;
- implementing a high level taxonomy of malware types as proposed in the present ontology conceptualization and leave further characterization to the interoperability with other cyber security formats.

As identified in the conceptualization, the malware indicators should be hashes (MD5 or SHA1), the filetype and the version. As it can be noticed in the ontology was not specified the name of the file. Naming a file can be not straightforward, where possible the name of the malware should be defined using the CARO Malware Naming Scheme[lxvi] which even if not updated is the most popular guideline among anti virus companies. In this respect, the need to a high level characterization arises to help the handler in quickly identify the threat. Moreover IODEF Fraud extension is too generic for a complete in-depth categorization of the sample and lacks of detailed information regarding behaviors and other characteristics that could be of help in response of the threat. The evaluation of a possible integration within IODEF data model of a high level categorization of malware is currently under discussion in the MILE mailing list. This effort could help in understanding the infection and can be neighborly in the foreseen integration with other cyber security exchange formats regarding malware.

Available interoperability with other cyber security exchange formats and specifications: unfortunately, at the time of writing there is no enumeration effort regarding malware samples, the Common Malware Enumeration (CME)[lxvii] initiative has been discontinued since 2007 and the efforts migrated into the DHS/DoD Software Assurance Forum Malware Working Group[lxviii]. Currently there are two different way to describe malware: Malware Metadata Exchange Format (MMDEF) and the Malware Attribute Enumeration and Characterization (MAEC). MMDEF prioritizes malware and creates high level indicators for exchange while MAEC is a standardized language and format based on pattern and attributes. In particular MAEC is tied to the Make Security Measurable (MSM) standards as CCE, CVE, CWE, etc. and enables

correlation and integration. The IODEF extension to support structured cyber security information already foresees the extension with the MAEC information. While this is an important step in the direction of enhancing an IODEF document with detailed malware information, most of the feeds that were analyzed do not make use of these formats and make use of proprietary scoring system as the Virus Total[lxix] so a further integration should be analyzed.

5.2.5 System

The system cluster identifies the appliance or the application targeted in the incident. IODEF offers several classes where these information are fully covered. The System, Service and Software classes are those in charge to cover these information regarding platform, version, application and patches.

Available interoperability with other cyber security exchange formats and specifications: several structured information standards are available and can extend the current IODEF classes such as the Common Platform Enumeration (CPE) that creates a naming schema for IT systems. Moreover common configuration issues can be listed using for Common Configuration Enumeration (CCE) while event logs can be exchange with Common Event Expression (CEE)[lxx].

6 Initial Requirements for an Application Scenario

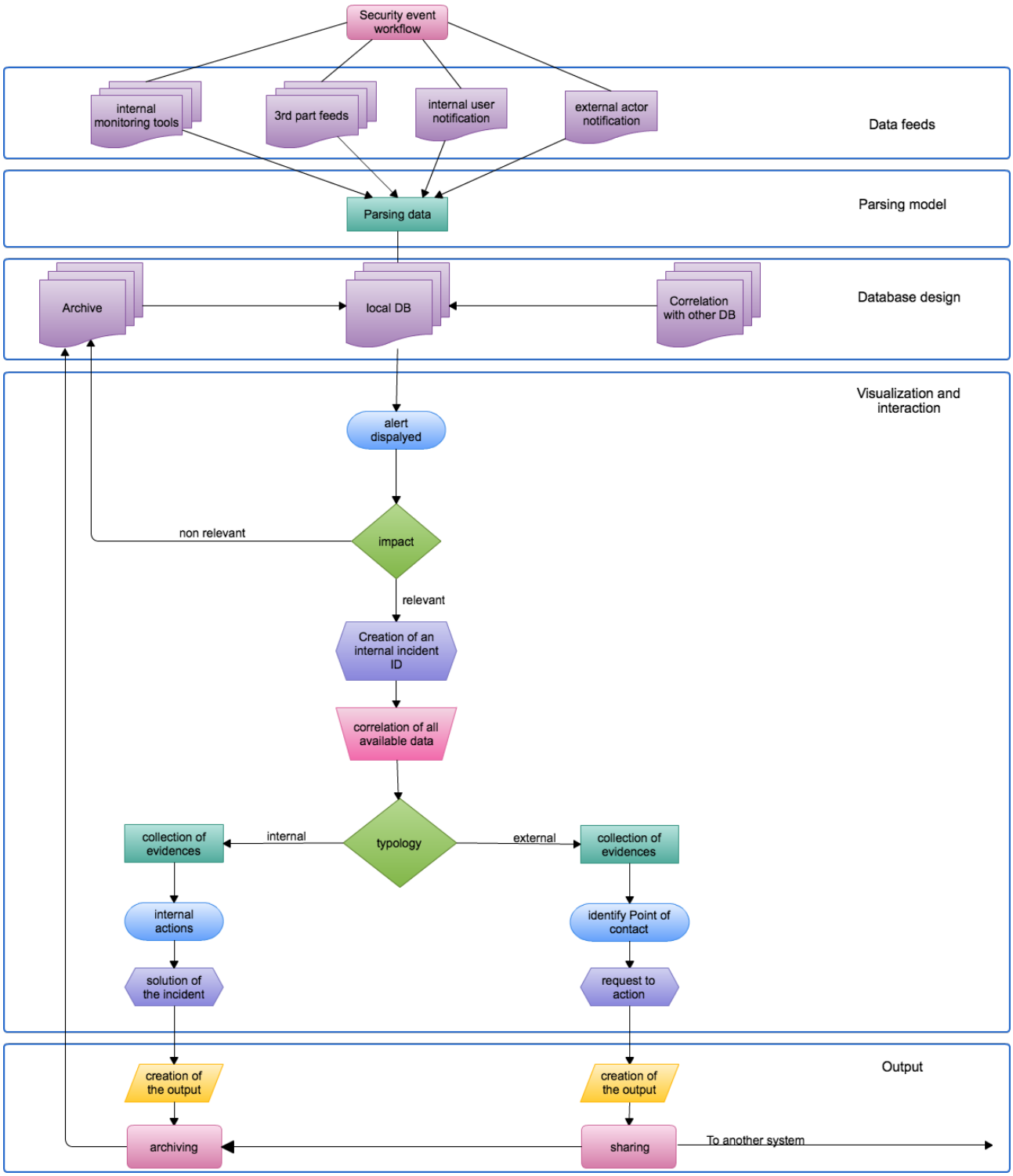
In the previous chapters it was first analyzed a corpus of publicly available data feeds in order to understand which kind of data could be used by a response team, then it was proposed a first ontology conceptualization of the data and a parsing implementation using IODEF data model was suggested. These phases were important to define metadata and content.Now some organizational variables regarding the intermediaries and the formats involved in the system will be depicted. As it was underlined at the beginning the goal of this research is to propose a comprehensive framework for response team to exchange information regarding security incidents and also pave the road to a standardization in the cyber security information exchange. In order to pursue this objective it will be here provided an overview of the other information needed to develop such a system. The goal is to cover all the potential requirements of a system that will be deployed in the next phases of this research.

6.1 A Possible Implementation

Data feeds can be external or internal. Moreover information can be received via electronic incident notification from other response teams or generated from the team members upon notification via other channels such for example telephone or other inputs. Electronic incident notification can be received via various formats and are most likely to be structured data related to a single incident. The notifications are sent by a response team or other entity and can include and extent all the clusters previously exposed. External data providers can be clearing house or publicly available services, both as malware domain lists or phishing domain repository, etc.. They have several provider specific metadata, can supply several different data feeds in various formats and can deliver different type of content related to several incidents or suspicious activities. The flow of the data should be the following: data is collected from external /internal feed and it is normalized using IODEF data model. Once collected the data is stored in the database and correlated with provided information regarding related malware samples (MAEC) or attack patterns (CAPEC), logs (CEE), platforms (CPE), vulnerabilities (CVE) also integrated with available vulnerability disclosure using CVRF and all other possible cyber security information. At this phase the data should also be correlated with other organizational related indicators regarding the supervised constituency, specific policy restriction and preferred output models to which the data can be subjected. Once the input data is parsed using the taxonomy expressed in the previous chapter is then first matched with the receiving constituency data. This

mechanism should allow to highlight only the data relevant to the response team and in future implementations could also allow to target alerting to a single team member. The team member can analyze the data using various clusters: addresses, malware, attack, system and incident metadata or utilize the different metadata for advanced or geolocated views. Eventual data related to other cyber security information can be correlated to the relevant indicators. Data from different inputs can be aggregated using these views and one or multiple instances can be selected showing all the information. Once created a view, the team member should be able to create a new incident document using the collected information. This document could be integrated by other team members findings using other tools or adding information related to the incident. The document could be then stored in the local database for archiving purposes or further investigation or prepared to be sent to an involved actor. Incident notification can be originated by team members collecting data from other sources, processing a notification that can not be automatically parsed, compiling a form using the fields that belong to the taxonomy previously exposed. Based on the previous chapter the main structure of the system should be based on the ontology implementation proposed before but should also contain interoperability functions with widely adopted proprietary tools such as Mandiant IOC, AlienVault Open Threat Exchange[lxxi], Verizon Enterprise Risk and Incident Sharing (VERIS) framework and application and other major adopted tools. Based on the information contained in the list of all possible teams, the document should be then formatted using the preferred format of the different receiver and information not complying to the sender/receiver privacy policy should be omitted/displayed due to reported specifications. In the following table a basic process overview is depicted in order to give the reader an idea of the flow of the information.

Figure 16 - A possible implementation: process overview.



6.2 Initial Requirements

In this chapter the previous ontology conceptualization and implementation will be matched

with the organizational information to define the base requirements for the implementation just depicted and that could allow response teams to use the data parsed and exchange this information. As it was underlined in the introduction, the proposed framework is willing to collect information from available data feeds. The idea is to allow the handler to have a complete picture of the incident and therefore permit a detailed information exchange. After the normalization phase of the incoming data using the ontology implementation, all information regarding incident should be aggregated primarily based on the address indicators. This is due to the fact that only the data that involves the constituency of the response team is relevant to be processed by that specific team. In order to define which data is relevant and how should be processed a set of more organizational related indicators should be applied and implemented in the framework. These indicators will be of help on three levels:

- identify the most interesting information for the response team based on the constituency perimeter;
- in case of involvement of other response teams identify which response team is in charge, if/which organizational requirements or restriction policy could be applied to data exchange;
- identify which formats and tools can be used in the exchange .

6.2.1 Organizational Information

In order to enable these actions the information regarding content and metadata should be integrated with a first set of all possible organizational information regarding data providers. As it was possible to understand in an incident exchange data providers are not only response teams or data sources but also other actors. Drafted on the Contact@role class present in IODEF, it should be possible to define this initial list of possible data providers using publicly available data and internal historical records, were possible:

- Admin - an administrative reference that reported or can be reported to for a specific host or network;
- Tech - technical contact that reported or can be reported to for a specific host or network;
- Response teams - response teams that could be involved in a event. This category

should enlist CSIRTs, CERTs and IRTs. Moreover due to frequent interactions with ISP and other organizations this list should be also contain network operation centers[lxxii] (NOCs) and abuse teams.

- LEA - Law enforcement agencies. Every day more LEAs are part of incident response and this is why a initial registry of the organizations that are most frequently involved in the exchange should be also considered.
- Sensor - As expressed in RFC 5901 a sensor can be " A intrusion detection system, firewall, filter, email gateway, or human analyst". This category should enlist both internal and external data feeds as previously presented at the beginning of the third chapter.
- Vulnerability reporter - Vulnerability reporting and management are an important part not only in the triage of an incident but can be also part of the proactive activity of a response team. For this reason the role vulnerability reporter was introduced in the document metadata.

6.2.2 Service Information

Most of teams publish a document regard their expectation of service. This document is drafted around RFC 2350[lxxiii], other actors have a website which contains all this data even if not structured around the RFC. In case no data is available the information can be inserted by human data entry. It is important to populate such database in order to have a primary definition of all the expected service characteristics of the actors from which data can be received or to which it can be sent. The structure of this view can be expected to follow the RFC 2350 even for those entities which are not a response teams in terms of CERT,CSIRT or IRT:

- Accuracy: in order to have updated situation of the information regarding the possible actors there should be a table containing the data regarding the last update and the location of the information. This information should be controlled via periodical check for updates.
- Contact information: modeled on IODEF contact class elements and the fields present in RFC 2350. This view should contain all contact information regarding the different teams as for example mailing address, time zone, telephone and fax number, other telecommunication means, email, public keys and encryption, team members and

operating hours.

- Charter: although at the beginning it will be mostly populated by free text format, where possible an initial effort should be considered in translating the constituency of all known response teams in network addresses. This would enable the immediate mapping of the constituency network addresses with the addresses involved in the incidents. As for some services the declaration of the network addresses in charge are required upon subscription, here the information should be natively inserted allowing automatic correlation. Moreover for response teams which comply to RFC 2350 further information regarding mission statement, sponsorship or affiliation and authority can be useful in the definition of the terms of the information exchange.
- Policies: as for the previous cluster this kind of information will be unstructured at the beginning but in the long run having a directory with all information about operational activity and disclosure information can be helpful in face some organizational issues like content restriction policy and secure means of communication. Where not available as in the RFC form (Types of Incidents and Level of Support, Co-operation, Interaction and Disclosure of Information, Communication and Authentication), information can gradually inserted by the handler.
- Services - incident response: this view should provide the main characteristics regarding the operational expectations. In case of incident exchange with an unknown response team can be of help in understanding what information send and what to expect. The RFC fields are incident triage, incident coordination, incident resolution, proactive activities and should be used as example.
- Incident reporting forms: the specifications provided in this view will help the handler to understand how to send information to another response team or what to expect to receive. At the moment there are no clear guidelines regarding the reporting forms.There are efforts[lxxiv] [lxxv] to fulfill this gap but none is widely adopted. The main goal of this research is to map this situation and provide a solution that will minimize this gap. Hopefully wherever here is provided a free text format or a structured document with the use of the presented ontology the problem will be not in how is data presented to handler but how automatically is parsed via the presented model.

Where not available all these information can be inferred by the website, the point of this

approach is to set a baseline for further development. Considering to establish the fundamentals of a platform for aligning response teams exchange, collecting this data can be considered useful to set a first database where then start building a more automatic mechanism for update. Other information that should be provided in order to fully cover the information exchange chain should also include:

- The role of the response team: directly related to the constituency is the nature of the response team (national, governmental, educational, vendor, private company, LEA,etc.) this will allow the handler to immediately understand the range of the counterpart.
- The affiliation to a particular organization: such as FIRST, INOC-DBA, TERENA TF-CSIRT, CERT.org or other organizations. Until now cooperation between response teams has been basically due to mutual trust and affiliation to organization. In this regard giving the handler the chance to understand if the response team is part of a partner organization can favor the exchange.
- Catalog of restriction policy: not every information regarding an incident can be exchanged to all response teams. With the evolution of threats also the information exchanged is subjected to several restriction policy and national and international privacy laws. The information regarding the use of a particular restriction policy as for example Traffic light protocol[lxxvi] or IODEF restriction policy should be available in order to understand how and which information can be exchanged.

6.2.3 Data Feeds Translation and Supported Standards and Formats

While the above information were more organizational, in order to complete this initial requirement some more related views regarding how the data incoming and out-coming in the information exchange is needed. This completes the requirements where from one side we have actual data regarding an incident incoming/out-coming the system and on the other sides is matched with organizational information to understand the contents and formats that can be exchanged. In order to enable this, the following information should be provided:

- list of all single data feeds: here it is possible to understand how the single data feeds are provided to the response team. This view enables to see if they are internal or external and which are the main contents they provide. Each single feeds should be related to a data provider listed at the beginning of this chapter.

- the template of document provided: a catalog of the incoming templates should be implemented in order to enable a quick review of the format the data feed is exchanged with. This should also contains the values from the ontology conceptualization to which the data is parsed: for every incoming template a mapping with the internal data model should be provided.
- the translation maps: it should contained the data models of all possible standards supported and the relative mapping schemas. For example it should contained the map between a IODEF document and VERIS document and therefore allow a semi automatic translation of the information between these two standards.

7 The Conficker Example

The purpose of this research is to suggest a comprehensive answer to various issues at different level of the information exchange. In order to give the reader an example of applicability of the present proposed framework, a brief scenario of use regarding the Conficker worm will be here depicted. Conficker still represents a current threat in the cyber security scenario[lxxvii] and it is also the first example of cyber security joint efforts thanks to the work of the Conficker Working Group. It is not in the purpose of this thesis to deepen the analysis of the worm, this is extensively covered by comprehensive research papers as "Know Your Enemy: Containing Conficker" of Felix Leder and Tillmann Werner[lxxviii] , "Detecting Conficker in your Network" from Adi Kriegisch[lxxix] and various analysis by Phillip Porras, Hassen Saidi and Vinod Yegneswaran[lxxx]. Here the idea is to use available information regarding this worm and frame them in the current system showing how this could have been of help in facing this threat. In the Conficker Summary and Review[lxxxi], Dave Piscitello, Internet Corporation for Assigned Names and Numbers (ICANN) Senior Security Technologist underlines some weaknesses in the experience of the Conficker Working Group:

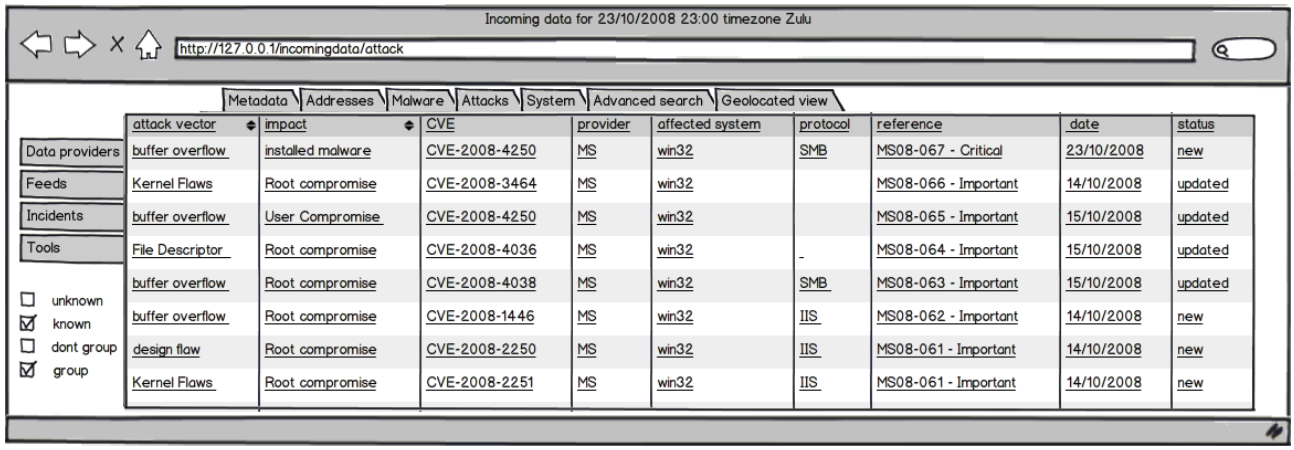
- *"Ad hoc collaborative response may not be scalable or sustainable;*
- *Informal communications may not be sufficient for all global incident response efforts, especially in situations where there is zero tolerance for error or omission;*
- *Maintaining consistency, completeness and accuracy of information during the course of a long incident response effort is challenging;*
- *Scaling trust is hard."*

Also in the report "Conficker Working Group: Lessons Learned" published by The Rendon Group[lxxxii] were noted as important items among others: "the need for collaborative infrastructure, information sharing, early warning and taxonomy". All these weaknesses could have been addressed with the implementation of this framework. Using the timeline[lxxxiii] of the infection the different chances of application will be presented in the following fashion. A screenshot of how this approach could have improved the incident exchange is presented.The screenshots make use of data acquired from the previously reported analysis or reports and try to sketch a possible alternative exchange which makes use of all that was previously exposed. For the purpose of this document only few episodes will be here provided in order resemble only the

main moments of the Conficker threat.

October 23, 2008 - Microsoft published the "Security Bulletin MS08-067"[lxxxiv] addressing a vulnerability that "could allow remote code execution if an affected system received a specially crafted RPC request". An hypothetical CERT-RS using the proposed framework could have parsed the vulnerability bulletin using the ontology. This could have allowed the information to be part of the structured information available to the handler. These information would have been placed in the attack and the system cluster.

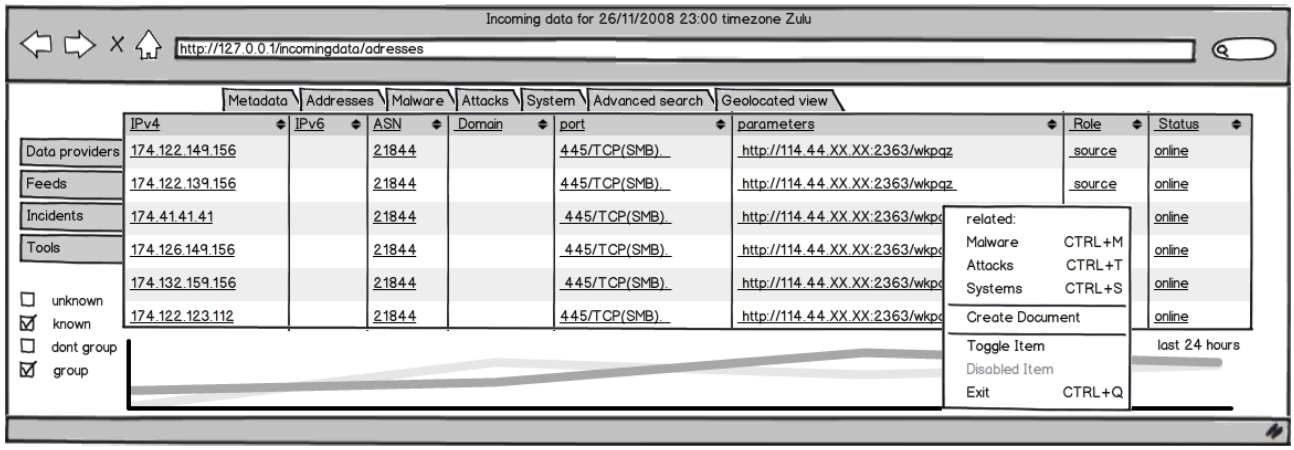
Figure 17 - Conficker - attack cluster: vulnerabilities index.



November 21, 2008 – Worm:Win32/Conficker.A initial release: various sensors started detecting the threat in their networks. Collecting feeds from all the different organizations and correlating information using the proposed ontology could have triggered an initial evaluation. Information regarding the addressees and the malware type and unique identifiers could have helped in understanding the initial recurring patterns. Assuming that the CERT-RS was in charge of all AS21844 addresses, all the incoming data feeds would have been normalized using IODEF ontology and then only those under the constituency AS would have bee shown. Then the handler could have started to see the same parameters on the same port. Considering that previously he had parsed the bulletin vulnerability from Microsoft a list of possible related vulnerabilities could have been correlated and a first analysis performed. In the following screenshot is depicted how the information about Conflicker.A could have been seen by the handler. The system would have shown only the addresses in the constituency of the handler, the targeted port and the requests performed by the infected hosts. With use of the other clusters the handler could have correlated information about the platform involved. Moreover if there were possible known vulnerabilities targeting the same ports and protocols these would could

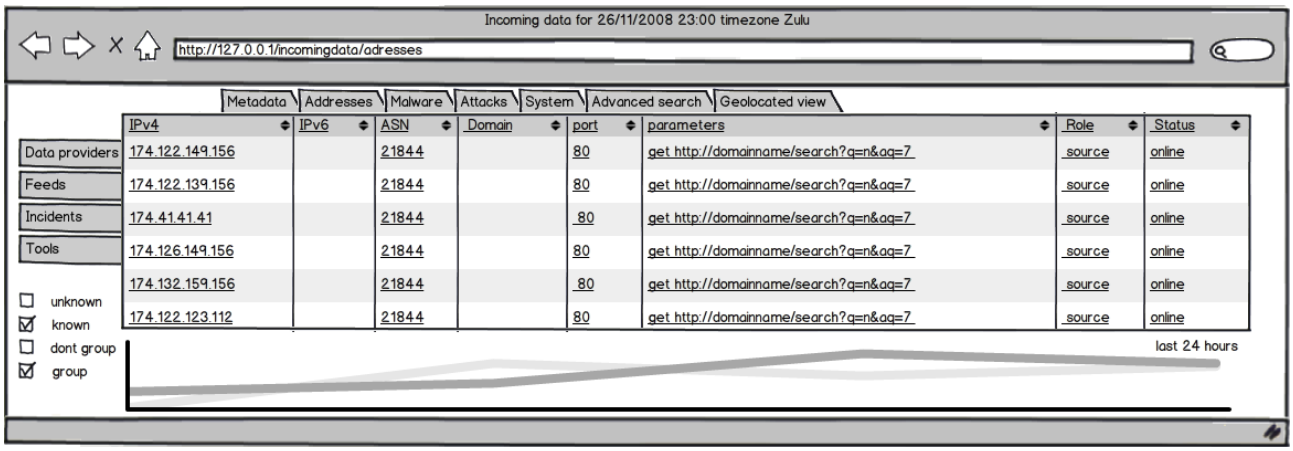
have been automatically shown.

Figure 18 - Conficker - address cluster: list of incoming data related to Conficker.A shellcode.



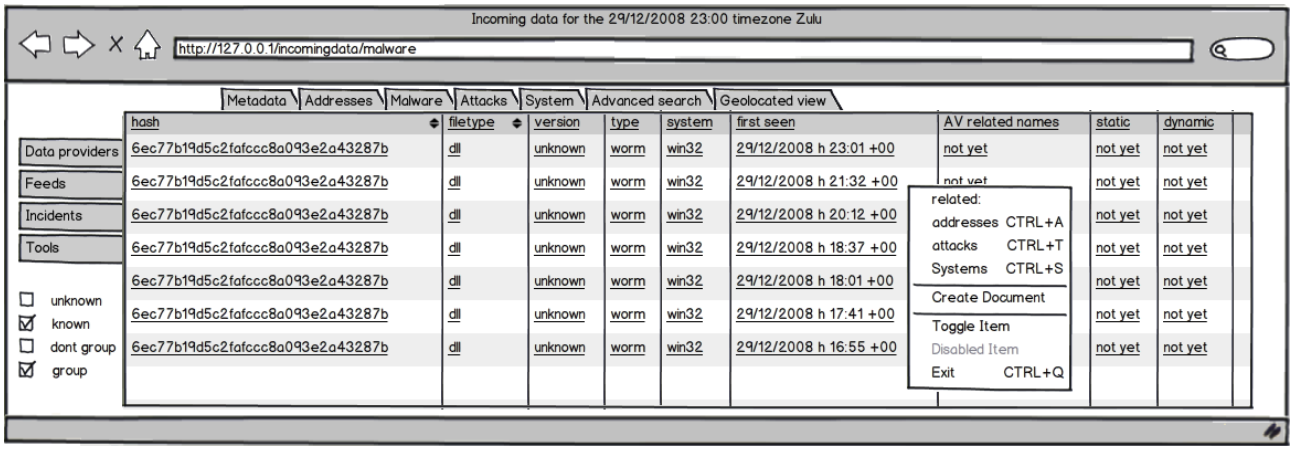
Considering that every infected host tries contact the automatically generated domain list, a peak of the same behavior could have been noted as in the following example. Due to the presence of all information regarding the response teams in charge of these domains, the handler could have started to exchange information incorporating in the document all the elements presented in the various clusters.

Figure 19 - Conficker - address cluster: Conficker.A attempts to connect to domains list.



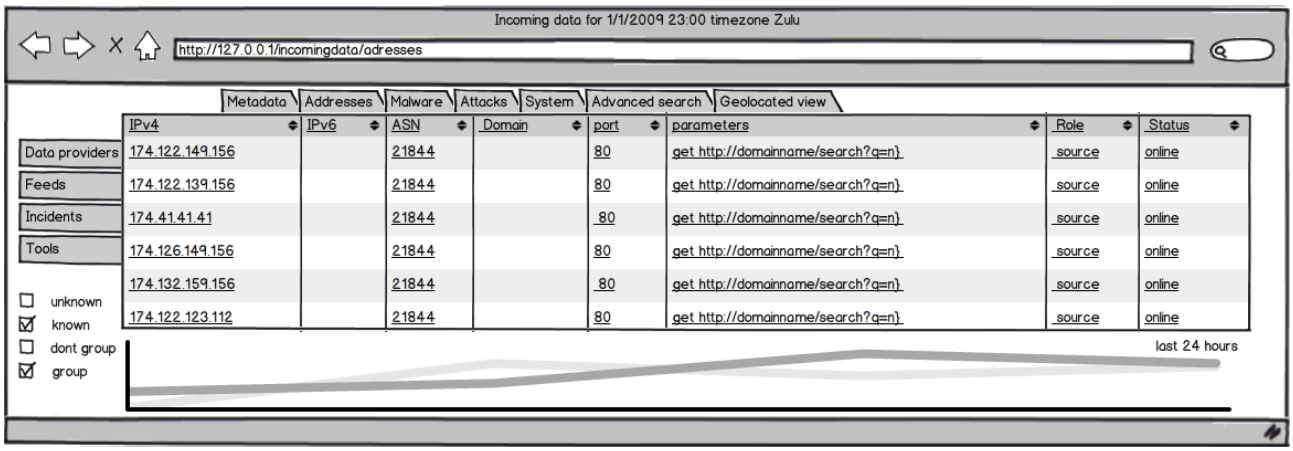
December 29, 2008 – Conficker.B released. Data arriving from various sensors could have been correlated using all clusters. Since the unique identifiers are different from the previous version, the malware would have been identified as a new threat. Previously gathered information could have been used to compare behaviors and findings from other response teams could have been integrated in the system. Although a comprehensive analysis of the worm was not present at that time, first high level indications regarding the most notable commonalities could have been noted.

Figure 20 - Conficker - malware cluster: Conficker.B detection.



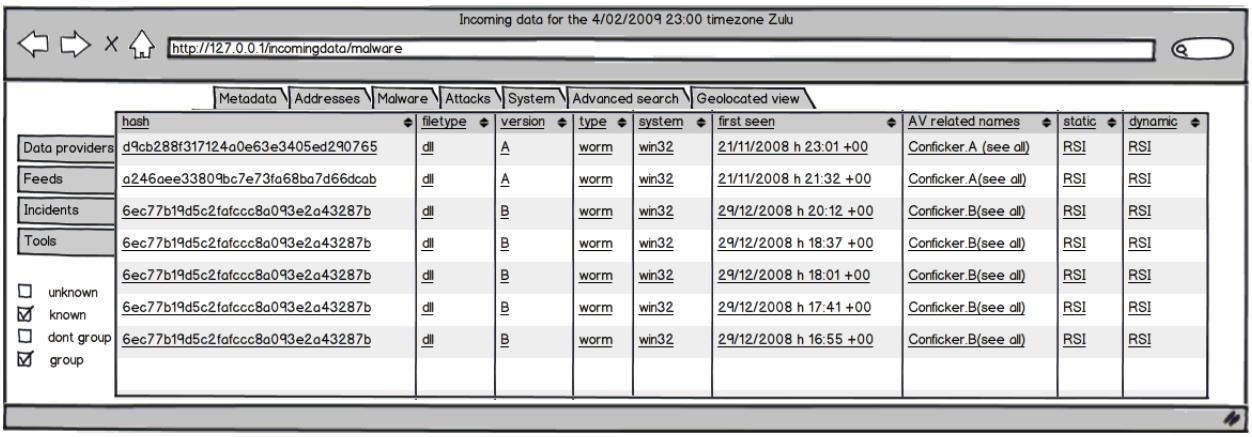
January 1, 2009 - Conficker.B payload activation date. Conficker.B starts to connect to the newly generated set of domains. The handler could have noted a change in the behavior of the requests, he could have then exchanged this information with the other parties involved or defined blacklists in order to block the traffic to these domains.

Figure 21 - Conficker - address cluster: Conficker.B attempts to connect to new domains.



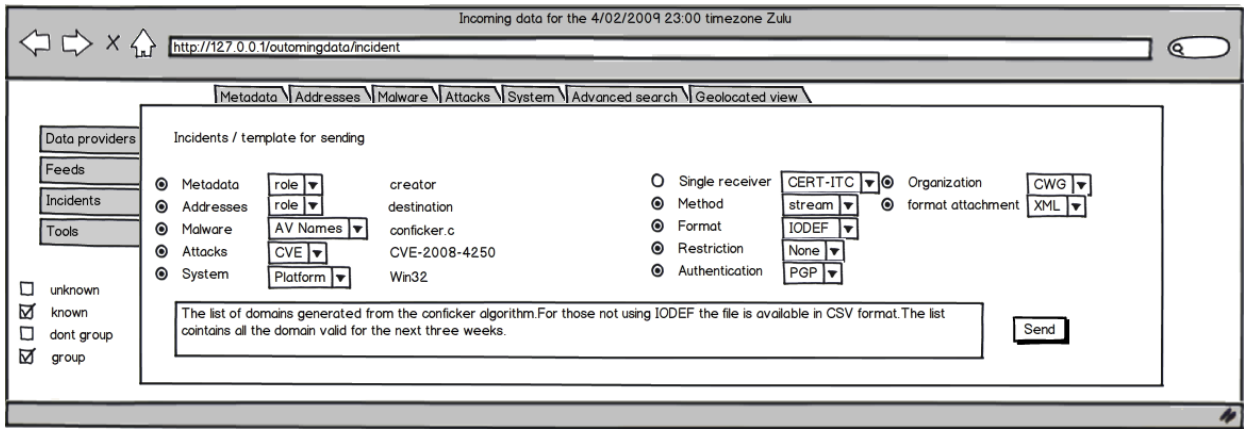
February 4, 2009 – SRI Conficker analysis published - The published analysis is imported in the system using the ontology implementation. Making use of proprietary data models available at the time or simply parsing the free text information, the data could have completed the already present high level information. In doing this the data from the external and internal feeds can have been correlated and the handler could have further tuned the filters in order to target the infection. In the following screenshot the view of the malware cluster after the parsing of the analysis is presented.

Figure 22 - Conficker - malware cluster: parsing RSI conficker analysis.



February 12, 2009 – Conficker working group initiated. CERT-RS could have suggested this approach and the WG could have adopted the present framework to exchange information. Data from various feeds could have been collected and parsed using the presented taxonomy. Since all information regarding teams involved in the system would have been present, data could have been exchanged automatically and securely due to the membership to the Conficker Working Group. Various formats could have been supported and the same content (in the screenshot below the list of the automatically generated domains) could have been exchanged using the preferred output of the receiver. Data could have been easily correlated. Moreover due to the presence of the restriction policy belonging to every specific team, information could have been omitted or presented in the single document exchange.

Figure 23 - Conficker - list of generated domains exchange with CWG members.



8 Final Consideration

This research shows how the current situation relative to cyber security incident exchange is variegated and proposes a way to start aligning it. As it can be noted the exchange of available contents is not paired both with a common language and an organizational answer that could enable response teams and other involved actors to correlate data, share and take actions in more comprehensive and rapid way. In the final recommendations of the ENISA report "Proactive Detection of Network Security Incidents" it is underlined: "Data providers are encouraged to adopt common formats for exchange of incidents while data consumers are encouraged to deploy correlation tools.". This proposal changes the perspective and aims to provide an answer to both needs with one solution. Various efforts have been made in the last ten years from various institutions and organizations in developing tools and specifications. The proposed framework is willing to use all these efforts in a more comprehensive way in order to give all actors involved a complete way to categorize and exchange incidents. The structure of the present research was therefore purposed to give the reader a full overview on what kind of solution could be implemented in the incident exchange. The will was to give an insight on the three level of the exchange: the content, the metadata related to the content and the organizational variables that should be addressed in a comprehensive solution.First the analysis of a corpus of publicly available data feeds was deployed in order to understand which kind of contents are accessible to the use of response teams.Then it was proposed a first taxonomy of incident data in which the same data was recollected in order to have a common exchanging platform where to start categorizing the content. Consequently a first implementation via a wide recognize standard and through the use of other available specifications and formats was proposed. In the end a feasible scenario of implementation covering the potential requirements of a system that could be deployed was presented including also an example based on Conficker data. As it can be noted the conclusions contained in this document are at the base for future implementations and developmenst which can be summarized in the following direction:

- provide an initial ontology for cyber security incident indicators,
- define the initial questions for the development of a taxonomy for attack vectors, attack impacts and malware categorization that could be widely recognized by the different actors of the security community as response teams, antivirus vendors and security researchers,

- propose the evolution of the IODEF data model in order to improve the adoption of this exchange format and the implementation of the current proposal,
- provide a preliminary study for the future software development of the proposed framework;
- pave the road for an initial standardization of the information exchange for cyber security incidents.

If from one side the landscape of security threats has reached a maturity and complexity that is everyday more difficult to frame, on the other side a comprehensive and adequate answer has not been developed yet. Security incidents can be transnational and highly complex and to face them the collaboration of various actors in a organized way is needed. The current proposal suggests a initial framework where to start aligning incident exchange among computer response teams and all actors involved such as anti-virus companies, reverse engineers, security vendors, data-providers and vulnerabilities reporters .Defining a system that covers such a complexity of data and actors and put the basis for a common ground can enhance the incident response both at local or constituency level and at a global level. Building the information society means exposing users to security threats but also creating an adequate protection. Shortening the gap between these two ends of the same continuum is an effort that is required nowadays. As exposed in the present research this effort requires to consider all the already available data, tools and specifications, provide a common language and consider also the organizational variables of all actors involved in order to maximize the result and provide a feasible and operative answer to global and variegated needs. Considering the different level of maturity and complexity of all entities involved, this could be considered an issue but it is also the opportunity to create a solution flexible enough that can fit all these needs in a leveraging way.

References

i Akamai, "State of the Internet report" http://www.akamai.com/stateoftheinternet/ Volume 4, Number 4, 4th Quarter, 2011, last retrieved 18th of may 2012

ii Forum of Incident Response and Security Teams (FIRST) members page, http://www.first.org/members/teams , last retrieved 18th of may 2012

iii Danyliw R., Meijer J., Demchenko Y. , "The Incident Object Description Exchange Format", RFC 5070, December 2007, https://datatracker.ietf.org/doc/rfc5070/ , last retrieved 18th of may 2012

iv Koivunen E. , "Effective Information Sharing for Incident Response Coordination: Reporting Network and Information Security Incidents and Requesting Assistance", Faculty of Electronics, Communications and Automation, Helsinki, 2010, p 36, 40, 45, 48-49, 54

v Dörges T., - 'Information Security Exchange Formats and Standards', Slides for the presentation held during FIRST 2009 Conference in Kyoto , 2009

vi Rutkowski T.," Significant Cybersecurity Developments: a global cybersecurity information exchange framework, plus Clouds, SmartGrid, and eHealth", Slides for the presentation held during ITU-T SG17 Tutorial Geneva, 2009

vii Mitre, "Making Security Measurable (MSM)", http://measurablesecurity.mitre.org/ , last retrieved 18th of may 2012

viii REN-ISAC, "SES - Security Event System" , http://www.ren-isac.net/ses/, last retrieved 18th of may 2012

ix Verizon, "Verizon Enterprise Risk and Incident Sharing (VERIS) framework and application", https://www2.icsalabs.com/veris/, last retrieved 18th of may 2012

x Abuse Helper, http://www.abusehelper.org/ , last retrieved 18th of may 2012

xi RTIR: RT for Incident Response, http://bestpractical.com/rtir/, last retrieved 18th of may 2012

xii Mandiant, "OpenIOC - Open Indicators of Compromise" , http://openioc.org/ , last retrieved 18th of may 2012

xiii H. Debar, B. Feinstein, "The Intrusion Detection Message Exchange Format (IDMEF), RFC 4765, March 2007, https://datatracker.ietf.org/doc/rfc4765/ , last retrieved 18th of may 2012

xiv x-arf - email format to report network abuse, http://www.x-arf.org/, last retrieved 18th of may 2012

xv Mitre, "Malware Attribute Enumeration and Characterization (MAEC)" , http://maec.mitre.org/ , last retrieved 18th of may 2012

xvi IEEE, "Malware Metadata Exchange Format (MMDEF) ", http://standards.ieee.org/develop/indconn/icsg/mmdef.html , last retrieved 18th of may 2012

xvii Mitre, "Common Attack Pattern Enumeration and Classification (CAPEC)", http://capec.mitre.org/ , last retrieved 18th of may 2012

xviii Mitre, "Common Vulnerabilities and Exposures (CVE)", http://cve.mitre.org/, last retrieved 18th of may 2012

xix Mitre, "Common Platform Enumeration (CPE)" , http://cpe.mitre.org/ , last retrieved 18th of may 2012

xx Mitre, "Common Configuration Enumeration (CCE)" http://cce.mitre.org/, last retrieved 18th of may 2012

xxi Mitre, "Common Vulnerability Scoring System (CVSS)" , http://www.first.org/cvss , last retrieved 18th of may 2012

xxii Mitre, "Common Weakness Scoring System (CWSS)", http://cwe.mitre.org/cwss/, last retrieved 18th of may 2012

References

xxiii Handover Interface and Service-Specific Details (SSD) for IP delivery, http://www.etsi.org/WebSite/Technologies/LawfulInterception.aspx , last retrieved 18th of may 2012

xxiv Electronic Discovery Reference Model - http://www.edrm.net/, last retrieved 18th of may 2012

xxv Howard J.D. , Longstaff T. A. "A Common Language for Computer Security Incidents" , Sandia National Laboratories, 1998, p 17

xxvi Gorzelak K., Grudziecki T., Jacewiz P., Jaroszewski P., Juszczyk L. and Kijewski P., " Proactive Detection of Network Security Incidents", ENISA, 2011, p 116 , last retrieved 18th of may 2012

xxvii In blue are underlined the sources that are also covered in the ENISA report " Proactive Detection of Network Security Incidents"

xxviii Hambridge S., Lunde A., "DON'T SPEW A Set of Guidelines for Mass Unsolicited Mailings and Postings (spam*)", RFC 2635, June 1999 https://datatracker.ietf.org/doc/rfc2635/ , last retrieved 18th of may 2012

xxixShirey R., "Internet Security Glossary, Version 2", RFC4949, August 2007, http://tools.ietf.org/html/rfc4949 , last retrieved 18th of may 2012

xxx Daley R., Millar T., Osorno M., "Operationalizing the Coordinated Incident Handling Model" , 2011 IEEE International Conference on Technologies for Homeland Security Proceeding, 2011 , p 287 - 294

xxxi Haldane A. G., "Towards a common financial language", Securities Industry and Financial Markets Association (SIFMA) "Building a Global Legal Entity Identifier Framework" Symposium, New York, March 2012, http://www.bis.org/review/r120315g.pdf , last retrieved 18th of may 2012

xxxii H. R. 3523 , Report No. 112–445 , "Cyber Intelligence Sharing and Protection Act" http://www.gpo.gov/fdsys/pkg/BILLS-112hr3523rh/pdf/BILLS-112hr3523rh.pdf , The Library of Congress, last retrieved 18th of may 2012

xxxiii Palma R., Hartmann J., Gómez-Pérez A., "Towards an Ontology Metadata Standard", 3rd European Semantic Web Conference, 2006

xxxiv IETF, "Request for Comments (RFC)", http://www.ietf.org/rfc.html , last retrieved 18th of may 2012

xxxv Pretty Good Privacy (PGP), http://www.symantec.com/theme.jsp?themeid=pgp , last retrieved 18th of may 2012

xxxvi Livingood J., O'Reirdan M., "Recommendations for the Remediation of Bots in ISP Networks", RFC 6561, March 2012, https://datatracker.ietf.org/doc/rfc6561/ , last retrieved 18th of may 2012

xxxvii Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifiers (URI): Generic Syntax", RFC 3986, January 2005`https://datatracker.ietf.org/doc/rfc3986/ , last retrieved 18th of may 2012

xxxviii Cotton M., Eggert L., Touch J., Westerlund M., Cheshire S. " Internet Assigned Numbers Authority (IANA) Procedures for the Management of the Service Name and Transport Protocol Port Number Registry" RFC 6335, August 2011, https://datatracker.ietf.org/doc/rfc6335/ , last retrieved 18th of may 2012

xxxixIdem reference viii

xl Idem reference iv

xli IANA IPv4 Address Space Registry, http://www.iana.org/assignments/ipv4-address-space/ipv4-address-space.xml , last retrieved 18th of may 2012

xlii Huston G., IPv4 Address Report , http://www.potaroo.net/tools/ipv4/index.html , last retrieved 18th of may 2012

xliii Paulauskas N. , Garsva E., "Computer System Attack Classification" Vilnius Gediminas Technical

References

University, Department of Computer Engineering, 2006

xliv Lowry Lough M., "A Taxonomy of Computer Attacks with Applications to Wireless" PhD thesis, Virginia PolyTechnic Institute, April 2001.

xlv Undercoffer J., Joshi A., Pinkston J., "Modeling computer attacks: An ontology for intrusion detection" , Proceedings of the 6th International Symposium on Recent Advances in Intrusion Detection (RAID'03), Lecture Notes in Computer Science, vol. 2820,2003, p 113–135

xlvi Simmons C., Shiva S., Dasgupta D., Wu Q., "AVOIDIT: A cyber attack taxonomy", Technical Report: CS-09-003, University of Memphis, August 2009

xlvii Alien Vault, "Malware: Exploring mutex objects", http://labs.alienvault.com/labs/index.php/2009/malware-exploring-mutex-objects/, last retrieved 18th of may 2012

xlviii Danyliw R., Meijer J., Demchenko Y. ,"Incident Object Description and Exchange Format Data Model and Extensible Markup Language (XML) Document Type Definition", April 2002, http://tools.ietf.org/id/draft-meijer-inch-iodef-00.txt, last retrieved 18th of may 2012

xlix Cain P., Jevans D. "Extensions to the IODEF-Document Class for Reporting Phishing", RFC5901, July 2010, https://datatracker.ietf.org/doc/rfc5901/ , last retrieved 18th of may 2012

l Mandiant, "OpenIOC - Open Indicators of Compromise" , http://openioc.org/ , last retrieved 18th of may 2012

li Mitre, "Cyber Observable eXpression (CybOX)", http://cybox.mitre.org/ , last retrieved 18th of may 2012

lii Arbor Networks supports IODEF

liii Bradner S.,"The Internet Standards Process -- Revision 3", RFC 2026, October 1996, https://datatracker.ietf.org/doc/rfc2026/ , last retrieved 18th of may 2012

liv An updated view of the status of these drafts can be found on the MILE website. Managed Incident Lightweight Exchange (Active WG) https://tools.ietf.org/wg/mile/ , last retrieved 18th of may 2012

lv IODEF-extension to support structured cybersecurity information https://tools.ietf.org/wg/mile/draft-ietf-mile-sci/ , last retrieved 18th of may 2012

lvi Expert Review for IODEF Extensions in IANA XML Registry https://datatracker.ietf.org/doc/draft-ietf-mile-iodef-xmlreg/, last retrieved 18th of may 2012

lvii Guidelines for Defining Extensions to IODEF https://datatracker.ietf.org/doc/draft-ietf-mile-template/ , last retrieved 18th of may 2012

lviii Klyne, G. and C. Newman, "Date and Time on the Internet: Timestamps", RFC 3339, July 2002. https://datatracker.ietf.org/doc/rfc3339/ , last retrieved 18th of may 2012

lix Icasi, "Common Vulnerability Reporting Framework (CVRF)" , http://www.icasi.org/cvrf , last retrieved 18th of may 2012

lx Microsoft, "Microsoft security updates and the Common Vulnerability Reporting Framework" , http://blogs.technet.com/b/msrc/archive/2012/05/17/microsoft-security-updates-and-the-common-vulnerability-reporting-framework.aspx , last retrieved 18th of may 2012

lxi Wysopal C., Christey S., "Responsible Vulnerability Disclosure Process ", https://datatracker.ietf.org/doc/draft-christey-wysopal-vuln-disclosure/ , last retrieved 18th of may 2012

lxii The Open Source Vulnerability Database, http://www.osvdb.org/ , last retrieved 18th of may 2012

lxiii The Exploit Database, http://www.exploit-db.com/ , last retrieved 18th of may 2012

lxiv Mitre,"Common Weakness Enumeration (CWE)", http://cwe.mitre.org/ , last retrieved 18th of may 2012

References

lxv Microsoft, “Microsoft Security Response Center Security Bulletin Severity Rating System” November 2002 , http://www.microsoft.com/technet/security/bulletin/rating.mspx , last retrieved 18th of may 2012

lxvi Computer Antivirus Research Organization (CARO) “Naming Scheme”, http://www.caro.org/naming/scheme.html, last retrieved 18th of may 2012

lxvii Mitre, “Common Malware Enumeration (CME)”, http://cme.mitre.org/ , last retrieved 18th of may 2012

lxviii DHS/DoD Software Assurance Forum Malware Working Group, https://buildsecurityin.us-cert.gov/swa/malact.html , last retrieved 18th of may 2012

lxix Virus Total, http://www.virustotal.com/ , last retrieved 18th of may 2012

lxx Mitre, “Common Event Expression (CEE)”, http://cee.mitre.org/ , last retrieved 18th of may 2012

lxxi AlienVault Open Threat Exchange (AV-OTX), http://www.alienvault.com/alienvault-labs/open-threat-exchange/ , last retrieved 18th of may 2012

lxxii Packet Clearing House, “INOC-DBA” http://www.pch.net/inoc-dba/, last retrieved 18th of may 2012

lxxiii Brownlee N., Guttman E., “Expectations for Computer Security Incident Response”, RFC 2350, June 1998 http://www.ietf.org/rfc/rfc2350.txt, last retrieved 18th of may 2012

lxxiv Cichonski P, Millar T., Grance T., Scarfone K., “Computer security Incident Handling Guide” , Special Publication 800-61 Revision 2 (Draft), National Institute of Standards and Technology, January 2012

lxxv CERT-EU, “Security White Paper 2011-002 - CERT-EU Services – Fundamentals” , Version 1.0 - 19 October 2011, http://cert.europa.eu/static/WhitePapers/CERT-EU-SWP_11_002_v2_1.pdf , last retrieved 18th of may 2012

lxxvi US CERT, “Traffic Light Protocol (TLP) Matrix and Frequently Asked Questions” http://www.us-cert.gov/tlp/, last retrieved 18th of may 2012

lxxvii Conficker Working Group, “Infection Tracking”, http://www.confickerworkinggroup.org/wiki/pmwiki.php/ANY/InfectionTracking , last retrieved 18th of may 2012

lxxviii Leder F., Werner T., “Know Your Enemy: Containing Conficker”, Honeynet Project, 2009

lxxix Kriegisch, A., “Detecting Conficker in your Network”, National Computer Emergency Response Team of Austria, 2009

lxxx Porras P., Saidi H., Yegneswaran V., “An Analysis of Conficker’s Logic and Rendezvous Points” SRI,2009, http: //mtc.sri.com/Conficker/ , last retrieved 18th of may 2012

lxxxi Piscitello D., “Conficker Summary and Review”, http://www.icann.org/en/news/announcements/announcement-11may10-en.htm , last retrieved 18th of may 2012

lxxxii The Rendon Group, “Conficker Working Group: Lessons Learned” , http://www.confickerworkinggroup.org/wiki/uploads/Conficker_Working_Group_Lessons_Learned_17_June_2010_final.pdf, last retrieved 18th of may 2012

lxxxiii Conficker Working Group, “Infection Timeline”, http://www.confickerworkinggroup.org/wiki/pmwiki.php/ANY/Timeline , last retrieved 18th of may 2012

lxxxiv Microsoft, “Security Bulletin MS08-067 – Critical” , http://technet.microsoft.com/en-us/security/bulletin/ms08-067, last retrieved 18th of may 2012