Taxonomy Model for Cyber Threat Intelligence Information Exchange Technologies

Eric W. Burger
Michael D. Goodman
Georgetown University
Washington, DC, USA
eburger@cs.georgetown.edu
mdg65@georgetown.edu

Panos Kampanakis
Cisco Systems, USA
panosk@cisco.com

Kevin A. Zhu
University of California
Los Angeles, CA, USA
kevinazhu@ucla.edu

ABSTRACT

The cyber threat intelligence information exchange ecosystem is a holistic approach to the automated sharing of threat intelligence. For automation to succeed, it must handle tomorrow's attacks, not just yesterday's. There are numerous ontologies that attempt to enable the sharing of cyber threats, such as OpenIOC, STIX, and IODEF. To date, most ontologies are based on various use cases. Ontology developers collect threat indicators that through experience seem to be useful for exchange. This approach is pragmatic and offers a collection of useful threat indicators in real-world scenarios. However, such a selection method is episodic. What is useful today may not be useful tomorrow. What we consider to be chaff or too hard to share today might become a critically important piece of information. Therefore, in addition to use case-based ontology, ontologies need to be based on first principles.

In this document we propose taxonomy for classifying threat-sharing technologies. The purpose of this taxonomy is to classify existing technologies using an agnostic framework, identify gaps in existing technologies, and explain their differences from a scientific perspective. We are currently working on a thesaurus that will describe, compare, and classify detailed cyber security terms. This paper focuses on the classification of the ontologies themselves.

Categories and Subject Descriptors

K.6.5 [**Management Of Computing And Information Systems**]: Security and Protection
H.3.5 [**Information Storage and Retrieval**]: Data Sharing
D.3.0 [**Programming Languages**]: Standards

General Terms

Security, Standardization, Languages, and Management.

Keywords

Information Sharing; Taxonomy; Ontology

1. INTRODUCTION

Security vulnerabilities and breaches occur at an alarming rate with no signs of slowing down. On an almost daily basis educational, financial, private and government institutions have been hacked and personally identifiable information (PII), intellectual property, and proprietary information stolen and used by various threat actors for monetary, personal or political gains. There is no single profile on threat actors. These actors can have many different motivations and are not bound by any specific tool or tools to accomplish their goals. In addition to threat actor profiles, there are different stakeholders in organizations including CEO's, CTO's, CISO's, Information Systems Security Managers, System/Network Administrators, System Architects, and System Users. Each of these stakeholders has a different role and consumes different forms or types of threat data in order to perform their duties and heighten the security posture of their organization. Different organizations and agencies have different structures and different individuals with a need to know the threat data. This creates a need for addressing security concerns and vetting of users and authorization to compartmentalized data, as well as formatting threat information for different uses.

As an example of different yet similar systems, the Department of Homeland Security (DHS) and the Department of Defense (DoD) both have vulnerability management systems that categorize and encapsulate threat data vulnerabilities cataloged in their own threat management systems. In addition there are different stakeholders within these organizations who need to know different pieces of the overall information or data. For example, the Information Security Vulnerability Management System (ISVM) [5] is a public threat data and remediation database that references the MITRE Common Vulnerabilities and Exposures (CVE) [14] for vulnerability and exposure information for operating systems, hardware, software, etc. Operating on this information, a CEO may only care if the threat and exposure is relevant to their operation. All they want is a binary yes or no answer. They are not concerned with the particulars of how to remediate the threat. An Information System Security Officer, on the other hand, or an Information System Security Manager needs to know more specifics: Is the threat relevant to the current infrastructure or operation? Has the operation been secured from the threat? If not, when will the remediation take place and what are the operational risks? System administrators need to understand the underlying remediation and potential operational risks such as testing and insuring remediation does not pose any downtime or additional operational risks. The roles and concerns are essentially the same in the DoD/DISA Vulnerability Management System (VMS) that maps the same MITRE CVE's to DISA IAVA's to U.S. Defense Information Agency's (DISA) Information Assurance Vulnerability Alerts (IAVA's) in the DISA Vulnerability Management system [22].

The threat landscape itself is constantly changing with known old vulnerabilities and exploits, phishing attacks, zero day attacks, denial of service, and other attacks. The characteristics and signa-

WISCS'14, November 3, 2014, Scottsdale, AZ, USA
Copyright 2014 ACM 978-1-4503-3151-7/14/11…$15.00
http://dx.doi.org/10.1145/2663876.2663883

tures of these attacks differ as to the carefully crafted responses offered by organizations, vendors and the community. The ultimate goal is to protect and secure consumer and organizational data from the loss of their assets, property, PII, and other data.

A common need among heterogeneous organizations and entities is to share different types of threat information about adversaries, targets and vulnerabilities. A major goal is to solve and distribute solutions to these threats in a timely manner and ultimately decrease the time between a zero day threat or vulnerability is discovered and when an action against that threat is initiated. A problem is there are multiple efforts underway for threat information sharing that use different data ontologies. These ontologies often overlap and do not offer a unified solution to the entire community. Typically they only address subsets of these communities. Thus, there could be duplications and gaps in the threat information sharing ontologies in different communities. This leads to a duplication of effort, and collaboration is not achieved or not achieved economically. There is a need for entities to have a common language and toolset to facilitate sharing. These different types of data, formats, roles, the need to know and privacy concerns creates a many-to-many Cartesian Product where a relationship that exists one day may not exist the next and new relationships are created using a mix of both old and new threat data. The taxonomy model presented here does not propose to solve these problems but provides an agnostic framework in which ontologies can be evaluated and assessed. Currently there are gaps and shortcomings in the Cyber Threat Intelligence Information Exchange Ecosystem. The goal is to identify these gaps and shortcomings for users, vendors, agencies and communities of interest to enable and provide scalable, robust and secure cyber threat information sharing.

After we present the taxonomy below, we examine two protocols and data representations, RID/IODEF and TAXII/STIX, using our model. RID and TAXII are the transport protocols for IODEF and STIX respectively. We also briefly examine YARA and NMSG.

2. Taxonomy Model for Cyber Threat Intelligence Sharing

Taxonomy is a classification into ordered categories. We propose a layered taxonomy model for cyber threat intelligence sharing technologies. We use a layered model, as opposed to a traditional hierarchical (taxa) model. This model, from a Computer Science perspective, better follows potential technology options and instantiations than a strict hierarchical model. This paper will not review the literature on separation of concerns or layered architectures. We refer interested readers to [24], [23], and [10].

The layers, shown in Figure 1, somewhat follow the ISO OSI protocol model. They are transport, session, indicators, intelligence, and 5W's (who, what, where, when, and why). The following sections describe these layers in more detail.

2.1 Transport

The transport layer is what moves the bytes representing the cyber threat intelligence between enterprises. There are three taxa in the transport layer. They are:

Pull Transport: Pull transport is a (potentially unbounded or unframed) stream of bytes requested by a client. An example of pull transport is HTTPS. A receiver polls for a message. The receiver knows they are receiving the message as they asked for it.

Push Transport: An asynchronous atomic message is a bounded message sent to a subscribing client. A sender packages a message and sends it to the receiver. The sender knows they are sending the message, but because of the asynchronous nature of message delivery, they do not know when or if the receiver will receive the message. Most asynchronous message protocols have provisions for positive notification of delivery. Examples of push transport are SMTP (Internet email), XMPP (Internet instant messaging), SIMPLE (Internet/3G multimedia instant messaging), and RSS feeds (Web/HTTP site update notification).

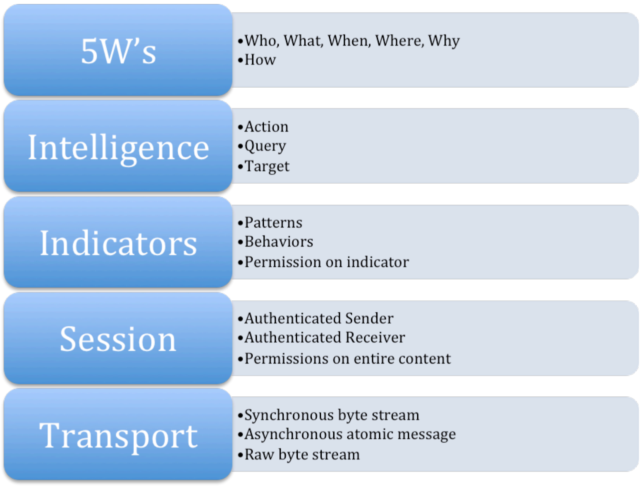


Figure 1 - Layered Model

Raw byte stream: One can always roll one's own message transport protocol.

Push transport can emulate a pull transport by having the sender 'request' the receiver to request the message from the sender.

2.2 Session

There are various services provided by the session layer. They include:

Authentication services: authenticating the sender as well as the receiver

Authorization services: given an authenticated identity, is the receiver trusted to receive and appropriately handle the information? Is the sender an authoritative source?

Permissions: Permissions on the entire content of the message

Authentication identifies the sender or receiver. However, authentication does not provide authorization. Just knowing who a sender or receiver is does not mean the sender is allowed to send the information, the information is authoritative, or that the receiver is allowed to receive it.

Some transports build in some level of authentication. For example, HTTPS identifies the server (host) by enabling the client to receive the server's X.509 certificate during TLS negotiation. HTTP can have the client identify the user using digest authentication. SMTP using S/MIME can identify the user or entity sending the message by signing the message and can cryptographically restrict access to the specified user or entity that is entitled to receive the message by encrypting the message with the recipient's public key (S/MIME or PGP) or via an application-level shared secret (password or key).

One can build authentication into the transport protocol. For example, if one has a VPN established, one can know with reasonable certainty the tunnel's endpoint, whether it is a network or specific host. For a number of use cases, such as point-to-point bilateral sharing, this is sufficient authentication for the sharing enterprises.

Other use cases do not require symmetric authentication. For example, take a CSIRT (Computer Security Incident Response Teams) that wants to publish information to the general public. In this case, only the CSIRT needs to present credentials. It is important for clients to be able to authenticate the CSIRT, as the CSIRT presumably is authorized to be the authoritative source of information. Since we are in the cyber threat environment, masquerading as a CSIRT can be a high value exercise. On the other hand, there is little technical value[1] in the CSIRT collecting strong identities of the clients. As the information is publically available, there is little technical value in encrypting the data between the server and the client, so long as there are ways of validating the information itself, such as through a cryptographic signature.

2.3 Indicators

The Indicators layer of the taxonomy is the first layer of the taxonomy that contains a cyber intelligence payload. There may be indicators that could require portion marking or encryption. The focus of much of the cyber threat intelligence exchange efforts to date is indicators. Indicators are patterns or behaviors that indicate, or show the likelihood and possibly predictability, of a cyber threat. Indicators can also be derived from incidents. If an incident occurs at some point in time there may be an observable pattern or behavior that indicates an incidence has occurred. Incidents can be considered as a set of indicators. As an example, in a distributed denial of service (DDoS) attack a service is unavailable. An Indicator of the attack could be a flood of packets from a certain region, identified by IP addresses, saturating perimeter routers.

Individual indicators may need permissions on them for privacy protection. For example, if an indicator contains PII, the sender may tag that particular instance of an indicator as protected. However, one can also have a need for cryptographically locking the indicator itself. For example, an indicator may include secret or non-forensic materials. Forwarding such information beyond particular personnel may expose methods or assets.

These permissions can be quite complex. Moreover, the permissions often depend more on the indicator's metadata than the indicator itself. That is, the permission is not dependent on the data element, but depends on data about the collection, provenance, ownership, source, use, and other factors not tied to the data element itself. This is a current area of research at Georgetown.

There is a distinction between an indicator and an incident:

- Indicator: indicates you may be under attack
- Incident: the attack that happened

In the case of an incident we detect the indicators of the attack, so we can act proactively on the attack at the Intelligence layer.

Observation of certain patterns does not necessarily imply one may be under attack; there can be false-positives. A simple example might be that if I receive a copy of the VBMania@MM malware in an email, it does not affect me, as I am on a Mac computer. Thus, the indicator pattern will be a false positive for me. However, forwarding the email to someone with a Windows machine may turn the indicator for the presence of VBMania@MM, into an incident, since Windows computers are compromised by this malware. One may even take this a further step: the infection of a Windows machine might be considered just an indicator, unless the virus terminates your anti-virus programs. However, even this could still be considered just an indicator (the machine is now running faster!), until the lack of anti-virus protection results in the installation of a backdoor that allows for massive exfiltration of one's data.

For the purposes of the taxonomy, we do not distinguish any 'incident' terms from 'indicators.' The Indicators layer of the taxonomy provides a separation between the action layer of the taxonomy (Intelligence) and the attribution layer of the taxonomy (5W's). In the taxonomy indicators need only have a secure transport and may operate independently from the Intelligence and 5W's layers. Also of note is that indicators are observable and are also a part of the 'H' for how in the 5W's Layer.

2.4 Intelligence

The Intelligence layer specifies action. It can be literally action, like "when you see this, do that." This is the way, for example, antivirus patterns work. "When you see a code snippet that looks like MyDoom, delete the file"; "When you see a http request to a banned domain, block the stream"; and "When you see an application open fifty TCP connections within 100ms, raise an alarm" are all examples of Intelligence.

The Intelligence layer also includes queries that are formulated from information accrued from the Indicator layer about a target or targets. For example, "What can you tell me about this IP address?" "What is known about this autonomous system?" "I am seeing this sort of behavior (indicators) – do you see anything like it?" are examples of queries.

Queries can be either synchronous or asynchronous. Early query systems were synchronous. However, often the data is not readily accessible or collected yet. This resulted in complex polling or, worse yet, no support at all. Asynchronous query mechanisms enable information security broker operations and, if not instant results, timely results.

2.5 5W's

The 5W's go back to at least the fourth century and are so basic to in information gathering, they are often mentioned in journalism, research and police investigations and constitute a formula for getting the complete story on a subject. This goes back centuries - Victorinus' Diagram is an example (see Figure 2). [26]

2.5.1 What are the 5W's

The 5W's are:

- Who is interested in this user or enterprise?
- What are they trying to do? What are they exploiting?
- When did it start?
- Where did it come from? Where does it go?
- Why are they attacking the enterprise or user?
- How does it accomplish their goals?

[1] For this analysis we ignore the potential marketing value of counting truly unique visitors, where they come from, and so on. Given the new distrust of government surveillance, not cryptographically identifying clients may be considered a feature, as some clients may not wish to register with the government simply to protect themselves from criminal bad guys.

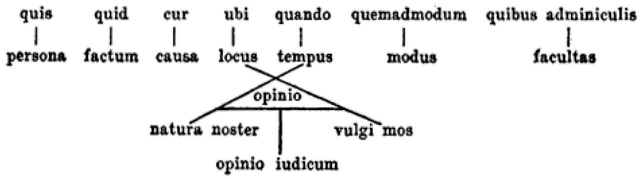


Figure 2 - Victorinus' Diagram of Cicero's Circumstances to Questions

The power of the taxonomy comes about when combining indicators with the 5W's. This is where one moves from indicators of varying, enterprise situation-specific importance to actionable intelligence. "When you see this behavior (indicators), you are a target and bad things may be happening" is one of the goals of automated intelligence exchange. Better situational awareness can be achieved by combining the indicators and 5W's of the taxonomy to promote more enhanced enterprise policies. An enterprise policy is a rule set that takes indicators and 5W's and results in actions the enterprise takes, either automatically, such as blocking a Web request to a known bad domain, or with manual steps, such as quarantining an email for later review.

To be clear, our taxonomy classifies intelligence sharing technologies, not the shared information. We use the 5W's to classify those sharing technologies that attempt to share information that we classify as 5W's.

The 5W's are above the action-oriented issues of the Intelligence layer. The traditional 5W's also known as the 5W's+H, or who, what, where, when, why and how. The Intelligence layer feeds information to the 5W's layer to provide attribution characteristics of who or what organization may be responsible for the threat. From the Indicator layer we ascertain threat patterns and behavior of threats as an indicator an attack has taken place or the potential an attack may take place. For example, a DDoS may have already taken place in one geographic location but may not have taken place in a different location. From these patterns or behaviors we analyze and formulate a plan of action to remediate the attack at the Intelligence layer, but we still may not know who the perpetrator of the attack may be. Post intelligence actions and through the use of queries and the target of the threat, we can start to formulate the 5W's. As an example, if the intelligence action is related to PII it may or may not be relevant to the target if the target is nuclear-related.

Understanding who is attacking and why can help the receiver of the information understand the urgency (or lack thereof) and likelihood (or not) of an attack on their enterprise, based on the indicators the enterprise sees.

With indicators we have confidence in what we know. We see this IP address or we do not. We see this snippet of code or we do not. Contrast this to what we think we know which is often the province of the 5W's. 5W's introduces a positive feedback process, as the 5W's can give us confidence in our interpretation of the indicators we are detecting. The action we take needs to be based on our understanding of what the indicator means in our current context, which we learn from an assessment of the 5W's.

The time value of indicators also complicates this. A false positive would be an enterprise taking action too soon, before validating that the indicator represents an actual incident. Conversely, an enterprise being too cautious and waiting too long for validation may get compromised.

For service providers (SPs), understanding the context of an attack is critical to deciding what to do. SPs need to be certain before taking an action that could affect millions of customers based on questionable indicators. However, knowing those indicators would set the service provider to take different, and quicker action than if they did not have the context provided by the 5W's layer or if they knew the indicators may be of low impact or value.

2.5.2 5W's

In the 5W's layer, *who* could be a person, organization or state for which the action of an attack could be attributed. In a cyber-attack, law enforcement officers are interested in *who* to arrest. In an enterprise, *who* could be used to incorporate into the Indicator layer as an indicator as a heuristic for what may be likely to happen. For example, it will help understand if the attack will be on economic assets, customer accounts, or designs, depending on whether the attacker (*who*) is doing blackmail, looking for cash, or a competing enterprise or state actor.

The 5W's give the big picture: above the layer of the individual incident or intelligence, the 5W's correlate multiple incidents, some that may seem unrelated. Besides mitigating current attacks, the 5W's can give clues to future targets and attack vectors.

The consumer of the information is one method to disambiguate the 5W's from intelligence. For prevention inside an enterprise, the enterprise's use of 5W's is similar to an indicator. The enterprise uses the 5W's information to make a decision.

The 5W's are particularly important for non-enterprise users. For example, law enforcement can apply non-cyber mechanisms to remediate the attack, such as arresting and convicting the perpetrators.

Understanding *Why*, one of the 5W's, can be important for the education of the enterprise. For example, if one makes widgets and someone wants the widgets, one may need to protect the data needed to make one's widgets. Knowing the motives or motivations can help in prevention, detection or correlating with other indicators to find the end goal of the attack and gain more intelligence. *Why* can be used as a heuristic and fed back to the Indicator layer and provide more threat information for the Intelligence layer as an action.

The more intelligence one has, the better notion of the 5W's you can gather. The better one understands the 5W's, the more confidence you can have in the actions (actionable intelligence) you take.

There are different kinds of confidence, some of which come from direct sharing, others come from corroborating evidence. For example, one may receive an initial assertion that a particular IP address is bad. There may be corroboration from, for example, SpamHaus, that the IP address is suspicious. However, you may also know the IP address is from, for example, a hosting provider that may have good tenants, too. Thus, the IP address goes from being marked as always bad to sometimes bad, based on the enterprise's analysis.

3. Extant Technology Classification

In this section, we will examine some information sharing technologies and map them to the proposed taxonomy.

We start with the IETF's Real-Time Inter-network Defense (RID) [20] and MITRE's Trusted Automated eXchange of Indicator Information (TAXII) [19]. These are technologies that map to the Transport and Session layers in our taxonomy model. They transport and create secure sessions for the interchange of cyber threat information objects. Incident Object Description and Exchange Format (IODEF) [4] and Structured Threat Information eXpression (STIX) [18] are examples of technologies at the Indicators, Intelligence, and 5W's layers of our taxonomy model. RID is used in combination with IODEF while TAXII uses STIX for cyber threat information sharing.

3.1 RID

RID is a transport for IODEF. RID is an Internet Engineering Task Force (IETF) standard defined in RFC6545. CSIRTs and SPs are the primary users of RID and IODEF.

RID fits into the taxonomy layer model on the Transport and Session layers as shown in Figure 3 below and is responsible for the transport of IODEF and IODEF-SCI extensions. Note, this considers the combination of RID and RID over HTTP/TLS. Here green indicates the technology is at that layer, while yellow indicates parts of the technology are in that layer.

5W's
Intelligence
Indicators (RID-dependent IODEF)
Session (RID)
Transport (RID over HTTP/TLS)

Figure 3 - Model for RID

3.1.1 RID Overview

RID provides secure and trusted transport sessions for the sharing of incidents and incident response information. In order to provide the most secure connection, RID is primarily used over HTTPS/TLS (RFC 6546) [28], encrypting the IODEF payload in transport. RID is derived from the IODEF data model and inherits data types defined in that model. RID prohibits internal or external DTD subsets. RID also provides session trust and affinity between vetted partners and subscribers. Privacy between partners and Service Level Agreements are enforced through the RIDPolicy class. RID provides secure sessions, while HTTP/TLS provides transport encryption. RFC6546 specifies RID may not use HTTP redirects; only TLS Version 1.1 (RFC2616) is allowed; both client and server must authenticate to one another using X.509 certificates; and RID systems must not respond to HTTP requests other than the HTTP response containing the RID response corresponding to the RID message. In theory, RID can carry payloads other than IODEF. In fact, Section 5.6.1 of RFC6545 describes how to register non-IODEF schemas and Section 12 establishes an IANA table for such. However, RID requires IODEF and a number of IODEF elements for proper protocol operation. Thus, RID is inseparable from IODEF, while the converse is not necessarily true.

RID defines a direct peer topology and a mesh peer topology. In a direct peer topology, a bilateral network peering relationship is defined for the members of a consortium. A mesh peer topology could also be used.

RID provides enhanced security at the Transport and Session Layers of our taxonomy, since authentication, authorization and trust are designed within the protocol using inherited IODEF data objects. RID requires trust relationships be established between all peers and provides an additional security layer. The advantages of RID are that it provides an extra layer of transport and session security using IODEF data objects in addition to HTTP/S and it provides authorization, authentication and trust relationships between CSIRTS and SP's. The disadvantages of RID include the following. First, the peer-to-peer communication topology does not scale for large networks. When nodes are added or deleted, nodes need to be either removed or added and appropriately whitelisted at every node in the network. Thus, RID peer and mesh topologies are better suited to smaller user communities. Also, RID is tightly coupled with IODEF, as RID requires IODEF data elements. Otherwise, messages will be rejected. More on this can be found in Section 3.1.3.

3.1.2 RID over HTTP/TLS

RFC6546 defines the transport protocol for RID over HTTP/TLS. As stated by RFC6546, RID describes some of the limitations of the use of HTTP as a protocol. RID does not expect immediate responses or document retrieval, and as such does not leverage HTTP semantics. Documents themselves are referenced using the HTTP POST method and are defined at the root level with Media and Content types of text/xml. The transport and session of RID can only be initiated to IP addresses of the machine that sent the request. For security reasons RID systems should not return HTTP 3xx redirection response codes. In real deployments, firewalls and other devices may use Network Address Translation (NAT). As such, IP addresses and ports could change, rendering such IP address whitelisting unusable or problematic. In order to not confuse RID with normal HTTPS port 443 traffic, RID uses port 4590.

Because security is a key requirement for information sharing, and as an IETF protocol the presumption is RID will be deployed on the open Internet. RID over HTTP/TLS mandates strong encryption and certificate policies. Specifically, RID mandates mutual, full path validation of X.509 certificates and non-NULL cypher suites. Nodes must be whitelisted at the RID client and server.

3.1.3 RID Proper

RID data types are derived from IODEF. RID requires a number of IODEF elements to be present, depending on the RID message type. RID/IODEF somewhat follows network-layering principles. For example, the RIDPolicy element describes the policy and destination information for the entire IODEF payload in the RID message. As separable XML, the concept is the RID XML elements would be in plain text, and the IODEF payloads could be encrypted and XML Digital Signatures provided depending upon the case as described in RFC6545.

As examples of the tight binding between RID and IODEF, the RID Request message requires the presence of at least eight IODEF data elements. The required IODEF classes for a RID Request message include: Incident class, Impact class, Confidence Class, System class, Expectation class, Event class, Record class, and the RecordItem class. IODEF timestamps are also required for RID.

The RID Result message requires IODEF data elements: Incident class, Impact class, Confidence class, System class, History class, Event class, and Record class and RecordItem class.

The RID Query messages require the IODEF data elements: Incident class, Impact class, Confidence Class, System class, History class, Event class, Record class, and RecordItem class. RID requires IODEF timestamps for the RID message and RID requires IODEF timestamps for the Request, Response, Report, Query messages. Missing any of the required IODEF classes from RID messages would be a violation to RID and cause errors.

RID requires at the minimum signatures on all IODEF payloads. For RID messages that would normally not have a data element, a time stamp is required to appear. Thus any RID payload can be authenticated and the possibility of a malicious attack that removes all contents from a RID message is eliminated.

What particularly ensures the tight binding between RID and IODEF is RID stores the routing history for the RID message in the IODEF EventData class, not in a RID class. This layer violation means that RID, although intended to be independent of IODEF, is in fact very dependent on IODEF.

The concept of separate encryption for the IODEF payload and RID Policy routing is useful. If the entire message were encrypted, intermediate nodes would need to decrypt the entire message to route the message. By separating encryption of the RID information from the IODEF payload the endpoint can do end-to-end encryption of the message using, for example, PGP, and also encrypt the routing information for the intermediate node. The intermediate node need only decrypt the routing information, not the end-to-end encrypted IODEF payload.

3.1.4 ROLIE

Resource-Oriented Lightweight Indicator Exchange (ROLIE) [8] is at the Transport and Session layers of our taxonomy using HTTP/S with the MIME media type of Application/Atom+XML. ROLIE is compatible with RID over HTTP/TLS, using existing HTTP message codes and IODEF data types and extensions. Figure 4 shows our classification for ROLIE.

A goal of ROLIE is to share incident and indicator information to a broad audience when there is no requirement for synchronous messages. This is different than the point to point and event driven model of RID/IODEF. For example, a government agency could publish information on the open Internet, as the incident and indicator objects are presented as web-based addressable resources. This is a resource-oriented approach and expands the number of stakeholders from just the CSIRTs in the IODEF community.

5W’s
Intelligence
Indicators
Session (ROLIE)
Transport (ROLIE over AtomPub)

Figure 4 – Model for ROLIE

One advantage of RID over REST versus HTTP/TLS is that notifications can be pushed to clients in a standard way. In addition, the ROLIE draft outlines how one could do service discovery using Atom service documents. The ROLIE work was not finished and has expired. One open issue includes how to do transport authentication.

3.1.5 XEP-0268 XMPP Incident Handling

XEP-0268 [9] describes transporting IODEF documents over the XMPP protocol. However, the XMPP Standards Foundation has deferred this work. It would be at the Transport and Session layers of our taxonomy.

3.2 TAXII

Trusted Automated eXchange of Indicator Information (TAXII) is at the Transport and Session layers of the taxonomy model as shown in Figure 5. TAXII uses HTTP or HTTP over TLS. TAXII defines a set of services and message exchanges for the exchange of actionable cyber threat information. This information can include IP addresses, email, domain names, and malware to discovered vulnerabilities and defensive courses of action between trusted partners [12]. TAXII is a community driven effort led by the U.S. Department of Homeland Security enabling secure threat information sharing.

TAXII uses network protocols such as HTTP/S to provide transport security and supports threat sharing models, including hub-and-spoke, peer-to-peer and publish-subscribe. TAXII messages do not convey authentication credential information. TAXII relies upon network protocols and network-level encryption to protect the entire TAXII Message in transit. TAXII is not tied to a specific payload, including STIX. The TAXII core components, services and message exchanges are defined separately from the implementation details of these components.

The TAXII specifications are agnostic in terms of content handling, content storage and access control mechanisms. Access control protections are implemented by content producers and not by TAXII.

The advantages of TAXII are a) different architectural models which scale to producers and consumers (hub-and-spoke etc.); b) implementation can support some, all or none of defined TAXII Services; and c) TAXII supports standard HTTP/S protocol and XML schemas and formats. The disadvantages of TAXII are as follows. TAXII does not provide producers and consumers separate methods for authentication, authorization and access to parts of cyber threat information that may be private, such as IP addresses. Producers provide these methods, as there is no standard way to configure it. Likewise, TAXII, while leveraging existing network protocols, does not provide additional authentication mechanisms or security within TAXII to enhance secure transport and authentication. Finally, producer data feeds cannot be edited. New information must be re-issued under an updated timestamp. This could possibly delay the availability of critical updated cyber threat information.

5W’s
Intelligence
Indicators
Session (TAXII)
Transport (HTTP, HTTPS/TLS)

Figure 5 - Model for TAXII

3.2.1 TAXII Overview

The Department of Homeland Security (DHS) office of Cybersecurity and Communications sponsors TAXII as a part of a cyber-threat information sharing system. MITRE has copyrighted the TAXII Specifications to ensure orderly change control. STIX is one example of data exchanged over TAXII. Since different communities may have varying security requirements either HTTP or HTTPS semantics may be used with TAXII depending on the authentication and encryption mechanisms, the security requirements, and the sensitivity of the threat indicators being transported.

3.2.2 TAXII HTTP/S Messages for TAXII Exchanges

The TAXII HTTP Protocol Binding Specification defines the transport of TAXII over HTTP and HTTP/S. TAXII can be configured to use the HTTP/1.1 (RFC2616) specification; however since TAXII Messages rely upon protocols such HTTP/S for authentication and encryption the use of the HTTP protocol should be avoided. The HTTP/S protocol can utilize different types of authentication such as X.509 certificates or userid and passwords.

3.2.3 TAXII Proper

TAXII defines push, pull, peer-to-peer, and combinations of transport sharing models. In the strict implementation of the push model, an authenticated and possibly authoritative source pushes threat information to subscribers of the threat information. In the pull sharing model, a subscriber can pull threat information from a source. The peer-to-peer sharing model allows the sharing and dissemination of threat information among multiple clients.

TAXII defines core services. Each service is optional and services can be combined in conjunction with any of the above sharing models. The Discovery service provides a way to discover what services are supported by some entity and how a client may interact with the service. The Feed Management service provides a way to learn and request subscriptions to data feeds. The Inbox service pushes threat content to defined inboxes. It could be sent via email or through a file. The Polling service provides a service where content can be requested using a pull model. The Query service provides a service where a request to a database or human gets a request and the results returned. In the same manner that sharing models can be combined, so can services be combined to accommodate different sharing models. Since services are optional, only applicable services may be implemented for a given deployment.

3.2.4 TAXII Discussion

TAXII authentication information is not conveyed in TAXII Messages. It relies on encryption at the network level and protocols to protect the messages in transit. Stakeholders may have different requirements for authentication and encryption. Enterprises are able to select which technology best supports their mission needs. Unfortunately, such a model requires considerable out-of-band setup to make operational.

TAXII presupposes an existing sharing agreement of some sort. On the Session Layer of our taxonomy model, users or organizational entities are authenticated as consumers/producers, senders/receivers, *etc*. The producers (senders) are entities that produce threat indicator information. Depending on the role, subscribers could have differing Service Level Agreements (SLA). Such SLAs could provide access to varying kinds of threat data. For example, in a community one organization may be interested in phishing attacks, others may be interested in malware, and yet others both. Some subscribers may opt for secure email (push). Others may opt for a listing and selection option (pull). TAXII provides information exchanges (or hubs) accommodating either of these options or a hybrid of services to consumers. Since there may be different or collections of communities of interest, there are gateways that connect hubs of common interest and distribute information, depending on methodology, via spokes to the end user or consumer. Some producers of cyber threat information indicators may share the usage of common gateways or use the same information aggregator.

Different communities of interest may be more interested in community-specific data formats. For example, there may be networking communities whose interest may be intrusion detection using SNORT or malware researchers using YARA for regular expression matching and analysis [12]. Both Snort and YARA can be expressed, encrypted, transported, and routed to the appropriate community gateway and to the end user using a specific service (push, pull, and hybrid) and SLA using HTTPS as the secure message exchange (see below for more on Snort and YARA).

Currently TAXII predominantly uses STIX as the cyber threat information binding being exchanged between subscribers and consumers in the STIX XML Schema.

3.3 IODEF

3.3.1 IODEF Proper

IODEF maps to the Indicators layer in our taxonomy model (see Figure 6). IODEF is an IETF standard format for representing computer security information commonly exchanged between CSIRTs and SPs and is at our taxonomy's Indicators and Intelligence layers. RFC5070 [4] describes IODEF.

The IODEF data model provides an XML representation for the sharing of threat information that is commonly exchanged by CSIRTs in relation to computer security incidents and events. Referencing RFC5070, some of the considerations in the design of the IODEF data model are a) the data model serves as a transport format; b) there is an assumption of a widespread agreement regarding the definition for an incident, therefore the data model does not attempt to dictate implementation; c) the intention of IODEF is as a framework extended for commonly exchanged incident information; d) IODEF balances free-form content and automated processing of incident information; and e) attempts were made to insure complimentary standardization to other security relevant data representations.

5W's
Intelligence (IODEF Queries)
Indicators (IODEF)
Session (RID)
Transport (RID over HTTPS/TLS)

Figure 6 - Model for IODEF

Although mostly an Indicator language, IODEF, in conjunction with RID, can be used as a query language at the Intelligence layer.

Currently, the APWG, APWG eCrime, XMPP Operators, Blackthorn and CIF communities use IODEF. The APWG (Anti-phishing Working Group) has extended IODEF in RFC 5901 to send phishing URLs. IODEF is not an adequate solution for packet captures and data types that are undefined [2]. However, the IETF has other solutions, such as IPFIX [3], to address this issue.

The advantages of IODEF are that it is a standard from the IETF and not the product of any one government. It is mature – the IETF first published IODEF in 2007. Moreover, IODEF can describe incidents in a granular manner. A disadvantage of IODEF is that it is fairly static, given the cycle times within the IETF.

3.3.2 IODEF-SCI

IODEF-SCI [27] extends IODEF to carry Intelligence information (see Figure 7). New classes include the Method class to map to AttackPatterns for CAPEC; Vulnerability for CVE or CVRF or CCE; Weakness for CWE; PlatformID for CPE; and Score for CVSS, CWSS and CCSS. The EventData class maps to the extended class EventReport used for the embedded information type CEE. The AdditionalData class maps to extend classes Verification for OVAL and XCCDF and Remediation for CRE.

5W's
Intelligence (IODEF-SCI)
Indicators (IODEF)
Session (RID)
Transport (RID over HTTPS/TLS)

Figure 7 - Model for IODEF-SCI

3.4 STIX

Structured Threat Information Expression (STIX) conveys the structure and information of the Indicators, Intelligence, and 5W's layers of the taxonomy model transported in a TAXII session in XML format. MITRE copyrighted the STIX Specifications to insure change control. The Department of Homeland Security (DHS) office of Cybersecurity and Communications sponsors STIX.

Figure 8 shows how STIX fits into our taxonomy. The recent TAXII query proposal enables the use of STIX for queries in the Intelligence layer, as well [6].

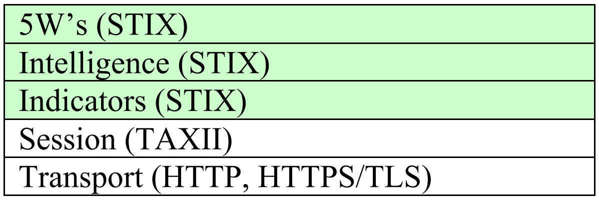


Figure 8 - Model for STIX

3.4.1 STIX Overview

The STIX Architecture is composed of eight core cyber threat concepts: Campaigns, Indicators, Observables, TTP (Tactics, Techniques and Procedures), Incidents, ThreatActors, ExploitTargets and Courses of Action.

Observables and Indicators are examples of constructs and patterns of STIX that map to the Indicators layer of the taxonomy model. Observables are the base construct in the STIX architecture. These are stateful or measurable events in computer or network operations, characterized by such attributes as names, hashes, services, *etc*. As a representation of Observables, STIX uses Cyber Observable eXpression (CybOX) [15]. CybOX is a schema for the specification, capture, and characterization of observable operational events. Indicators are comprised of specific Observable patterns in conjunction with contextual information intended to represent artifacts and/or behaviors of interest in a cyber-security context. New indicators are constantly being defined using communities of interest knowledge and best practices that are constantly evolving.

Incidents, TTPs, and Courses of Actions are examples of STIX constructs that exist on the taxonomy Intelligence layer [1]. Incidents are comprised of discrete instances of indicators along with the discovery of information pertaining to threat actors, exploit targets, and information gathered from the 5W's layer of the taxonomy model. TTP is the intelligence gathered on adversaries that are specific representations of the behavior, resources, targets, targeted victims, *etc*. This is the hub of all the threat intelligence information gathered in order to define specific courses of action in order to remediate the threat. CoursesofAction are the direct result of all the threat intelligence gathered in order to protect an asset or individual privacy data from the specific threat and corrective action taken by cyber threat management.

Campaigns, ThreatActors and ExploitTargets in the STIX architecture are examples of the 5W's layer of the taxonomy. Campaigns are characterized by the intent of a threat actor observed through sets of incidents and TTPs and attributions to the ThreatActors believed to be associated with the campaign. Other characteristics of the campaign are the confidence and reliability of the source of the campaign information, handling of the response taken or guidance, and activity taken in the response. STIX leverages community knowledge and best practices to define a new Campaign structure for representing Campaign information. ThreatActors are the Who in the 5W's layer. These are the characterizations of the malicious actor or actors; their motivations, intents, history, *etc*.; and the confidence in the identities of these actors. As in Campaigns, the content representing the ThreatActor is not structured but consists of text, so community knowledge and best practices allow for different or updated content to define new ThreatActor structures and represent different ThreatActor information. ExploitTargets are systems that can be exploited using weaknesses in the software or in the configuration of the software or the system. Such weaknesses can result in the compromise of the system causing loss of property, privacy and/or information. STIX uses identifiers such as Common Weakness Enumeration (CWE), Common Vulnerabilities and Exposures (CVE) and the Open Source Vulnerability Database (OSVDB).

3.4.2 STIX Discussion

STIX has core and common schemas, which provide the framework and common characteristics. It has more detailed schemas representing the attributes specific to the eight core cyber threat concept schemas. Within each of these eight component schemas, the specific schema can express information relevant to each concept. The component schemas can be used individually or in combination with one another, using whatever is relevant to a threat or threat area.

In addition to CybOX, STIX can embed XML namespaces such as CAPEC, MAEC, OVAL, CVRF, and CIQ. The Common Attack Pattern Enumeration and Classification (CAPEC) schema attributes characterize how cyber threats are executed and provide ways to defend against these threats. The Malware Attribution and Enumeration Characterization (MAEC) schema provides a standardized language about malware based on behaviors, artifacts and attack patterns [16]. The OVAL schema (Open Vulnerability and Assessment Language) uses the language standardizes the assessment process of representation of configuration information for testing; analyzing the machine state for patches, vulnerabilities, *etc*.; and reporting the results [17]. The Common Vulnerability Reporting Framework (CVRF) was derived from IODEF (Incident Object Description Exchange Format) and somewhat complements and somewhat competes with STIX [11]. CIQ (OASIS Customer Information Quality) are a set of XML based specifications for defining, representing, interoperating and managing "PARTY (Person or Organization) CENTRIC INFORMATION" [21].

Other Indicators that STIX could carry include IODEF in lieu of the IncidentType extension within STIX. Mandiant OpenIOC [13] extensions can be used in lieu of CybOX to express non-standard Indicator patterns at the Test_Mechanism extension point. In addition there are extensions for Snort and YARA rules. The XML specifications for CIQ representing information about individuals and organizations can be expressed in the STIX schema extension Identity construct. This can be used for the common information regarding information such as threat actors and victims.

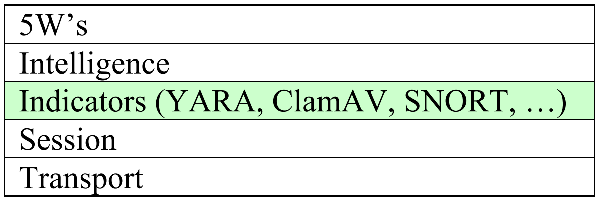
Non-XML STIX bindings could be developed, using, for example, JSON. The above bindings should support block level encryption that enable granular level access to different roles or individuals for an organization or agency and allow a non-repudiation level of security

Cyber threat intelligence information sharing is a global problem that requires global acceptance. Today, DHS and FS-ISAC are using STIX and other U.S.-based critical infrastructure sectors are working on implementing STIX. One issue facing STIX and TAXII is they are a product of the U.S. Government, which may hinder adoption by other countries that would share information but would not accept a U.S. Government standard. DHS plans to take TAXII and STIX forward for standardization in a global standards development organization. [25]

The advantages of STIX are that it is modular and can describe different indicators and fields and it can incorporate other standards efficiently. The disadvantages of STIX are that it is driven by a government organization and is very complex to implement. As with IODEF, if there were multiple representations of objects, a gateway would be required to translate between models. [12]

3.5 Other Technologies

YARA [30] (Figure 9) is a pure indicator layer technology that describes regular expression patterns and behavior. YARA is an engine and language for scanning files and memory blocks. When a rule matches a pattern, YARA presumes to classify the subject according to the rule's behavior. YARA can match various string formats like ASCII, UTF, and other encodings; case sensitivity (either sensitive or insensitive); *etc*. YARA can also parse on PERL regular expressions. YARA has Python and Ruby bindings.



5W's

Intelligence

Indicators (YARA, ClamAV, SNORT, …)

Session

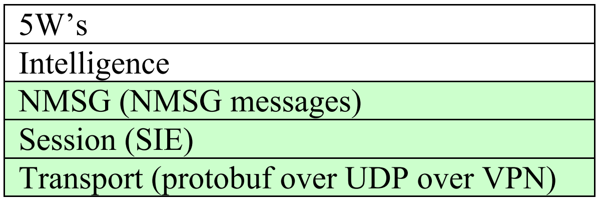
Transport

Figure 9 - Model for YARA, ClamAV, etc.

There are a number of technologies similar to YARA that are at the Indicators layer, such as ClamAV, PEiD, SNORT, Bro, and Unix/GNU regular expression parsing tools. ClamAV is open source and is popular for email filtering. There is a Python script that will convert ClamAV rules into YARA expressions. PEiD searches Windows portable executable files for strings, is open source but does not appear to be actively maintained. There is a Python script that will convert PEiD specifications into YARA expressions. SNORT and Bro are open source packet analyzers with intelligence to match rule sets and trigger actions when there are expression matches. These occur at both the Indicators and Intelligence layer of the taxonomy.

The Verizon Enterprise Risk and Incident Sharing (VERIS) [29] system is a characterization of cyber incidents after they have occurred, intended for strategic trend analysis and risk management and metrics framework.

Figure 10 described how NMSG [7] fits into the taxonomy. The DNS community is the primary user of NMSG. NMSG occurs at the Transport, Session and Indicators layers of the taxonomy. NMSG specifies a transport, leveraging Google Protocol Buffers (protobuf) as an encoding format and API. Using the protobuf model, NMSG defines a number of different payload containers for transport and storage. Data transmission is UDP broadcast over static virtual private networks (VPNs). Because of the use of UDP, NMSG is not compatible with IP fragmentation and has a proprietary header at the NMSG layer to fragment messages larger than the maximum IP datagram size of 64KB. Because of this, guidance for NMSG buffer size is 8KB. Session security is achieved via VPN provisioning. There are proprietary headers in different vendor defined message formats. NMSG can report on many indictors and is extensible to report on other indicator types.



5W's

Intelligence

NMSG (NMSG messages)

Session (SIE)

Transport (protobuf over UDP over VPN)

Figure 10 - Model for NMSG

4. CONCLUSIONS

The layered taxonomy model proposed in this work allows us to assess and analyze different Cyber Threat Intelligence Exchange Ontologies. Using the taxonomy we are able to decompose these ontologies and better analyze dependency and interoperability within the current cyber threat sharing communities.

Our taxonomy model abstracts the cyber threat intelligence secure information exchange into distinct layers. Each of the layers can be used to assess different cyber threat intelligence secure information exchanges.

Most often, the Transport and Session layers ride over HTTP/TLS; this provides the integrity, encryption and transport of the payload from and to different entities. On the Transport layer, TLS handles encryption of the byte stream, which can be synchronous or asynchronous. In this way the confidentiality and integrity of the raw data can be assured. Those systems that use HTTP/TLS encapsulate in the Transport layer some Session layer information. The Session layer is responsible for the authentication and authorization of users. Since there are wide ranges of stakeholders and combinations of consumers and producers, the Session Layer defines the way users are authenticated to the system and what threat data they are allowed to access.

The upper three layers (Indicators, Intelligence, and 5W's) are community-specific. The Indicators are the observable patterns, behaviors and events and the permissions on who is allowed to access which Indicator. Indicators will be different for stakeholders depending upon the characteristics of the threat. There are different Communities of Interest. Clients who may be more or less vulnerable to particular threats and may have a need to know, whereas there are other communities do not. Indicators can define a community or communities whose responsibility falls into that domain. Indicators may not necessarily be clear-cut from one stakeholder group to another stakeholder group. As an example a phishing exploit may eventually lead to exploiting some other hardware or software vulnerability. So, Indicators need to have the ability to be shared and automated in some fashion to expedite the analysis for the Intelligence layer.

After the Indicators have been parsed and analyzed to gain specific knowledge regarding the target exploit, intelligence is gathered and an action plan is formulated and aggregated at the Intelligence layer as an actionable item from the intelligence gathered. Intelligence at this layer is dependent on the analysis of the indicators at the Indicator layer. Some valid questions at this layer are: Does the exploit exist? Who or what is the target of the exploit? And, what is the resolution or antidote for the exploit? Resolutions and fixes need to occur as quickly as possible to facilitate fixing the vulnerability and preventing the spread of the attack. As intelligence is gathered regarding the target, methodology of the exploit, *etc*.; the information is used to put together information on the adversary and possible motivations of the adversary. That is the role of the 5W's layer.

Related work underway at Georgetown is a thesaurus of terms defined by the various ontologies mentioned in this paper. So far, we have integrated the STIX and IODEF definitions. While STIX provides a broader scope of terms that cover concepts such as Indicators, Courses of Action, and TTPs (Tactics, Techniques, and Procedures) in an overall STIX package, IODEF fills some of the gaps, such as more details for Contacts and descriptions about the victim(s) of the cybersecurity attack. In particular, IODEF focuses more on the Incident itself and categorizes terms revolving around the central concept of an incident. Because of this focus, many of the concepts of IODEF can be integrated within STIX's architecture. Although IODEF includes basic information about related concepts, such as Threat Actor and Campaign, as part of an incident, these ideas can be merged into STIX's other major concepts with more detail. Other general relationships include IODEF's Method match to STIX's TTP and the similar representations of start, end, and detection times.

Another avenue of further research is to explore whether the Traffic Light Protocol (TLP) is sufficiently rich to express the complex trust relationships in the cyber threat intelligence information exchange market. This relates to the issues around data element marking and metadata-based permissions discussed in Section 2.3.

5. ACKNOWLEDGMENTS

We would like to acknowledge the anonymous reviewers who provided valuable insights and suggestions for improving this paper. Members of the S^{2}ERC Georgetown CyberISE Project contributed industrial knowledge and experience. In particular, we would like to thank Merike Kaeo from IID and Shawn Wilson from Verisign for their work in the program.

This material is based upon work supported by the National Science Foundation under Grant No. 1362046 and the industry affiliates of the Security and Software Engineering Research Center (S^{2}ERC). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the S^{2}ERC affiliates. Support for this work includes funding from the S^{2}ERC double affiliate Symantec; S^{2}ERC affiliates AT&T, Check Point Software Technologies, Cisco Systems, Comcast, Edgewater Networks, and Verisign; and S^{2}ERC associate affiliate IID. Affiliates pay Georgetown University and the funds are used to cover the expenses of the study and related academic and research activities of the institution.

6. REFERENCES

[1] Barnum, S., *Standardizing Cyber Threat Intelligence Information with the Structured Threat Information eXpression (STIX™),* http://stix.mitre.org/about/documents/STIX_Whitepaper_v1.0.pdf, page 11. Accessed: July 16, 2014.

[2] Cain, P., *APWG Adventures In Information Sharing: Now and For the Future*, http://scap.nist.gov/events/2011/itsac/presentations/day2/Cain%20-%20Advenures%20in%20Info%20Sharing.pdf. Accessed: July 16, 2014.

[3] Claise, B., Trammell, B., Aitken, P., *Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information*, IETF RFC 7011, September 2013.

[4] Danyliw, R., Meijer, J., Demchenko, Y., *The Incident Object Description Exchange Format*, IETF RFC5070, December 2007.

[5] DHS, *DHS Sensitive Systems Policy Directive 4300A*, Version 8.0, DHS, March 14, 2011.

[6] Davidson, M. and Schmidt, C., *The TAXII Default Query Specification: Version 1.0*, MITRE, January 13, 2014.

[7] Farsight Security, https://github.com/farsightsec/nmsg, Accessed July 30, 2014.

[8] Field, J., *Resource-Oriented Lightweight Indicator Exchange*, IETF Internet Draft draft-field-mile-rolie-02, August 15, 2013, work in progress (expired).

[9] Hefczyc, A. *et al.*, XEP-0268: Incident Handling, XMPP Standards Foundation draft XEP-0268. May 29, 2012. Work in progress.

[10] Herzberg, D. and Marburger, A., "The use of layers and planes for architectural design of communication systems," in **Proceedings Fourth ISORC**, May 2001, pp. 235-242.

[11] ICASI, "The Common Vulnerability Reporting Framework," http://www.icasi.org/cvrf. Accessed: July 16, 2014

[12] Kampanakis, P., *Survey: Security Automation and Threat Information Sharing Options*, to appear in **IEEE Security and Privacy Magazine**, September/October 2014.

[13] MANDIANT, "OpenIOC," http://www.openioc.org/. Accessed: July 22, 2014.

[14] MITRE, "Common Vulnerabilities and Exposures," https://cve.mitre.org/. Accessed: July 23, 2014.

[15] MITRE, "Cyber Observable eXpression," http://cybox.mitre.org/, Accessed July 16, 2014.

[16] MITRE, "Malware Attribute Enumeration and Characterization," https://maec.mitre.org, Accessed: July 16, 2014.

[17] MITRE, "Open Vulnerability and Assessment Language," https://oval.mitre.org, Accessed: July 16, 2014.

[18] MITRE, "Structured Threat Information eXpression," http://stix.mitre.org/. Accessed: July 16, 2014.

[19] MITRE, "Trusted Automated eXchange of Indicator Information," https://taxii.mitre.org. Accessed: July 16, 2014.

[20] Moriarty, K., *Real-time Inter-network Defense* (RID), IETF RFC6545, April 2012. Accessed July 22, 2014.

[21] OASIS, OASIS Customer Information Quality (CIQ) Technical Committee, https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=ciq. Accessed: July 16, 2014.

[22] OUSD, "Information Assurance Vulnerability Management," http://www.prim.osd.mil/cap/iavm_req.html?p=1.1.1.1.3, Accessed: July 29, 2014.

[23] Parnas, D., "On the criteria to be used in decomposing systems into modules," in **Communications of the ACM**, v. 15 n. 12, December 1972, pp. 1053-1058.

[24] Savolainen, J. and Myllarniemi, V., "Layered architecture revisited — Comparison of research and practice," in **WICSA/ECSA 2009**, September 2009, pp. 317-320.

[25] Struse, R., *Release of STIX 1.0 and CybOX 2.0*, email dated 10 April 2013 to the MITRE STIX discussion list.

[26] http://en.wikipedia.org/wiki/Five_Ws, Accessed: July 23, 2014.

[27] Takahashi, T., Landfield, K, and Kadobayashi, Y, *An Incident Object Description Exchange Format (IODEF) Extension for Structured Cybersecurity Information*, IETF RFC 7203, April 2014.

[28] Trammel, B., Transport of Real-time Inter-network Defense (RID) Messages over HTTP/TLS, IETF RFC6546, April 2012. Accessed July 22, 2014.

[29] VERIZON, *Verizon Enterprise Risk And Incident Sharing Metrics Framework*, https://www.verizonenterprise.com/resources/whitepapers/wp_verizon-incident-sharing-metrics-framework_en_xg.pdf, Accessed: July 16, 2014

[30] YARA, http://plusvic.github.io/yara/. Accessed: July 16, 2014.