DIMS Operational Concept Description Documentation

Release 2.9.1

David Dittrich

Apr 28, 2018

Contents

Executive Summary

Since HSPD-7 was released in 2003, the Department of Homeland Security has had a core mission of working to protect the nation's critical infrastructure. In 2008, the *National Response Framework* was released, and a project to take tools developed by DHS Science and Technology for use in federal government networks and put them in the hands of *State, Local, Territorial, and Tribal (SLTT)* government entities – known as the *Public Regional Information Security Event Monitoring (PRISEM)* project – was initiated. *[Note1]* The intent of the *PRISEM* system was to combine standard security devices event log data using a commercial *Security Information Event Management (SIEM)* system, fed in part by event log data from the DHS-funded NetFlow based system (formerly known as *Einstein 1*), correlating these events using the SIEM to detect structural bot activity that has a high probability of being an infected computer. It used the *Collective Intelligence Framework (CIF)* database system to produce watchlists for real-time monitoring, as well as to provide historical attack context. A geographic front end provided a regional context to alerts in the system for at-a-glance situational awareness. The system allowed indicators of compromise (IOCs) to be used for both finding events that were missed in the past and/or watching for new events in the future.

DHS efforts with MITRE to develop information sharing mechanisms based on the *Structured Threat Information eXpression (STIX)* format have made de-classified *Indicators of Compromise (IOCs)* and *Observables* available to regional SLTT government entities, allowing them to confirmation involvement of threat actors of national interest. As this sharing of IOCs and linked Observables is extended laterally to similar regional collaborative efforts, national scope and visibility of the impact of widespread threats becomes possible.

The *Distributed Incident Management System (DIMS)* project is intended to do two things.

- First, to take this semi-automated sharing of structured threat information, building on the success of the PRISEM project and leveraging the portal system used by an existing community of operational security professionals known as *Ops-Trust*, and scale it to the next level. DIMS will take advantage of the open "message bus" architecture developed under PRISEM, features that support "identification of friend or foe," and the ability to integrate three data sources maintained by PRISEM (network flow history, event history, and attacker context history) to support the triage process, cross-organizational correlation of events, and anonymization to promote privacy-sensitive sharing of security event data. Working with the use cases defined by MITRE and PRISEM users, building the features necessary to simplify structured information sharing, and operationalizing these within these existing communities, will allow DIMS to fill existing gaps in capabilities and support existing missions that are slowed down today by many complicated, manual processes.
- The DIMS project also aims to establish a model open source framework and "scaffolding" that will promote the integration of open source computer security tools to provide a feature-rich, flexible, scalable, and affordable toolset for regional responce and information sharing efforts.

CHAPTER 1

Scope

1.1 Identification

This Operational Concept Description (version 2.9.1) describes the operational concepts for the Distributed Incident Management System (DIMS).

1.2 System overview

DIMS is funded by the Department of Homeland Security under contract HSHQDC- 13-C-B0013. For more information, see the document, "System Requirements and Concept of Operations for From Local to Gobal Awareness: A Distributed Incident Management System (DIMS)" referenced in Section *Referenced documents*.

The primary mission objectives for the DIMS system are operational in nature, focused on facilitating the exchange of operational intelligence and applying this intelligence to more efficiently respond and recover from cyber compromise. The secondary mission objectives are to create a framework in which tools to support the primary mission objectives can more quickly and easily be integrated and brought to bear against advancing techniques on the attacker side of the equation.

The DIMS project is intended to take this semi-automated sharing of structured threat information, building on the success of the Public Regional Information Security Event Monitoring (PRISEM) project *[Note1]* and leveraging the portal used by an existing community of operational security professionals known as Ops-Trust, *[Note2]* and scale it to the next level. The intent of this software project is to allow for near real-time sharing of critical alerts and structured threat information that will allow each contributing party to receive information, alerts and data, analyze the data, and respond appropriately and in a timely manner through one user-friendly web application.

Working with the use cases defined by MITRE and PRISEM users, building the features necessary to simplify structured information sharing, and operationalizing these within these existing communities, will allow DIMS to fill existing gaps in capabilities and support existing missions that are slowed down today by many complicated, manual processes.

The changes to existing systems consists of seamless integration of the three current systems into a single web application that enables each system to contribute to the data warehouse of information concerning threats, alerts, attacks and suspect or compromised user terminals within the infrastructure. Additionally, the integrated systems will be able

to share and retrieve data, visually observe alerts through color coded visual indicators, while retaining the existing functionality of the current system.

1.3 Document overview

The structure of this document has been adapted principally from MIL-STD-498 (see Section *Referenced documents*). Following this section are:

- Section *Referenced documents* lists related documents.
- Section *Current system or situation* describes the current PRISEM system, its sub-components, their capabilities and limitations, the existing user base, and support concept.
- Section *Justification for and nature of changes* describes the justifications for how the current system needs to change, why those changes are relevant, alternatives, and assumptions/contraints.
- Section *Concept for a new or modified system* describes the concept of a new and improved system and related issues.
- Section *Operational scenarios* provides operational scenarios that will drive requirements and the system's architectural design.
- Section *Notes* provides an alphabetical listing of acronyms and abbreviations used in this document.
- Section *License* includes the copyright and software license under which DIMS is being released.

CHAPTER 2

Referenced documents

1. DIMS System Requirements v 2.9.0
2. DIMS Test Plan v 2.9.1
3. DIMS Commercialization and Open Source Licensing Plan v 1.7.0
4. HSHQDC-13-C-B0013, "From Local to Gobal Awareness: A Distributed Incident Management System," Section C - Statement of Work
5. MIL-STD-498, Military Standard Software Development and Documentation, AMSC No. N7069, Dec. 1994.
6. Organization Design: A Sustainable and Self-Sufficient Model for Washington State's PRISEM Partnership, by Parker Montgomery, University of Washington Daniel J. Evans School of Public Affairs, March 2014.

Note: See also the Bibliography at the end of this document.

CHAPTER 3

Current system or situation

3.1 Background, objectives, and scope

Homeland Security Presidential Directive 7 (HSPD-7), *[Exe03]* released in 2003, set in motion a number of policy changes and created awareness of a new problem – the term for which is now ingrained into our lexicon: critical infrastructure protection. That document specifies the actions to be taken to identify, prioritize, and address the vulnerabilities to the systems and services that have relevance to the American way and quality of life. Local (city and county) government provides systems and services that maintain and improve the quality of life at the scale at which citizens identify most directly. eGovernment services allow citizens to pay bills, obtain a business license, communicate with elected officials, etc. Local government arguably maintains 85% of critical infrastructure, yet its protection is largely unaddressed. While the rest of the world is focused on nation-state hackers versus Supervisory Control and Data Acquisition (SCADA) systems or cardholder data breaches, local governments with their silos, poor budgets and priorities that change with the political wind are attempting to protect water purification, traffic management, public safety communications and a great many other services – the loss of which would arguably have an impact including loss of life. Local government is acknowledged in HSPD-7 as being integral to critical infrastructure protection, however no reasonable effort has been made to address cyber defenses on the local scale. Efforts to date to secure cyberspace have called for a comprehensive system that protects federal government agencies. A "comprehensive system" for securing the United States electronic infrastructure that does not include local governments, however, is not truly comprehensive.

Taking this all into consideration, two things become clear:

- Critical infrastructure dependencies on local government must be addressed
- Local governments need assistance with detective controls – security monitoring – and with response capabilities

Both these issues may be addressed by extending a concept that is common to corporate IT organizations into the local government sector: managed security service. Specifically, were a central location available for securely routing activity logs, firewall and IDS alerts, and other forms of information typically collected (but not analyzed) on networks, and then made available to local governments as near real-time alerts and a portal for situational awareness, the gaps between the criticality of the systems and services, and the degree to which critical infrastructure elements are being protected can be addressed.

The Public Regional Information Security Event Management (PRISEM) system was designed to address gaps in capabilities between federal and local government entities. PRISEM extends a concept common to corporate IT orga-

nizations – managed security services – into the local government sector. It enhances security oversight and controls and improves the ability to detect and respond early to threats against critical infrastructure. It moves beyond basic information sharing and creates an action-oriented alliance that leverages limited expertise across resource-constrained local government IT organizations. It creates a partnership between a top-tier research university, federal law enforcement fusion center, and private sector organizations. Its benefits will include increased security and compliance capabilities, increased productivity, improved performance, and lower costs for participants.

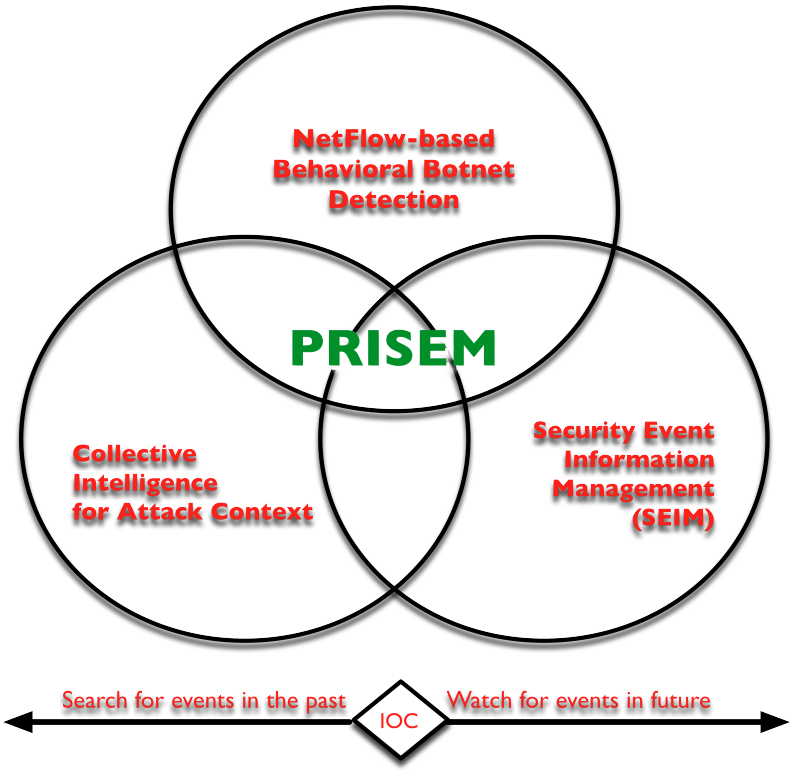

Fig. 3.1: PRISEM capabilities

The intent of the PRISEM system is to combine standard security devices event log data using a commercial Security Information Event Management (SIEM) system, fed in part by event log data from the DHS-funded NetFlow based system (formerly known as *Einstein 1*), correlating these events using the SIEM to detect structural bot activity that has a high probability of being an infected computer. It uses the Collective Intelligence Framework (*CIF*) database system to produce watchlists for real-time monitoring, as well as to provide historical attack context. A geographic front end provides a regional context to alerts in the system for at-a-glance situational awareness. The system now allows indicators of compromise (IOCs) to be used for both finding events that were missed in the past and/or watching for new events in the future. This is depicted in Figure *PRISEM capabilities*.

The primary mission of the PRISEM system is threefold:

- To **enhance the information security capabilities** of local government and address exposures to critical infrastructure, systems and services without significantly raising cost, by providing the means to obtain visibility into attacks against information technology resources;
- To **provide a method for reporting** cyber-security event or trend information in a consistent and automated fashion, for further evaluation by intelligence or law-enforcement communities in a manner that is respectful of national and international standards of individual privacy; and
- To **create an action-oriented operational setting** for the deployment of research-grade technologies that were

funded by the federal government, in order to evaluate their effectiveness and assist with their transition into commercial products.

In 2008 The Federal Emergency Mangement Agency, part of DHS, released the National Response Framework. *[Fed08]* The relationship building between hometown security and Homeland security began to form an enduring partnership. As part of its commitment to hometown security, "DHS has worked to get tools, information, and resources out of Washington, D.C. and into the hands of our federal, state, local, tribal and territorial law enforcement partners." *[Dep13]* The PRISEM project, initiated this same year, is an example of this effort to bring these resources to the SLTT government level. It has served this purpose, but to date only in the Puget Sound region.

Fast forward to February 2013. The President of the United States issues two new policies:

1. Executive Order 13636: Improving Critical Infrastructure Cybersecurity *[Exe13a]* and
2. Presidential Policy Directive 21: Critical Infrastructure Security and Resilience. *[Exe13b]*

These two documents (known as *EO 13636* and *PPD 21*) reflect the acknowledgement that:

- America's national security and economic prosperity are dependent upon the operation of critical infrastructure that is increasingly at risk to the effects of cyber attacks.
- The vast majority of U.S. critical infrastructure is owned and operated by the private sector and/or State, Local, Territorial, and Tribal (SLTT) government entities, not by the federal government.
- A strong partnership between the public and private sector, as well as between SLTT government entities in regions of the country, is crucial in reducing the risk to these vital systems.

3.2 Operational policies and constraints

3.2.1 Policies

- EO 13636 and PPD 21 provide guidance on how the federal government will work with private sector operators of critical infrastructure systems in order "to prepare for, prevent, mitigate, and respond to threats."
- Policies for each of the SLTT government and private sector entities participating in the PRISEM system, and the PRISEM participant agreement, have privacy impacts when sharing information outside the project.

3.2.2 Assumptions

- It is assumed that the Ops-Trust portal system will be easy enough to refactor to accommodate the required API for user interface enhancements that underlie the DIMS front-end.
- In addition, a successful application penetration test result (and remediation of critical security flaws that these tests may uncover) is a pre-requisite for the Ops-Trust stewards to allow the code to be released to the general public.
- It is assumed that the open source tools necessary to provide the full set of capabilities described here and in the DIMS DIMS System Requirements v 2.9.0 document, can be assembled in such a manner that they provide the necessary features in a coherent and integrated a manner.
- We assume that the stakeholders who have expressed an interest in providing requirements and beta-testing feedback will follow through. It will be important to have at least two groups (beyond the Ops-Trust community and US-CERT) perform some "live-fire" structured information sharing experiments in order to fully exercise the data sharing aspects of DIMS. It is hoped that an organization like NCFTA, who is already familiar with the Ops-Trust portal system, can facilitate development and testing of the specific information sharing features that are part of their daily business processes.

3.2.3 Constraints

- Data currently held in the PRISEM system cannot be shared with non-PRISEM members without the express permission of those whose data is held in the system. The DIMS team is operating under an NDA with the City of Seattle for access to *the City's data* in the PRISEM system for development purposes. Anonymization features described in this document are intended to facilitate sharing within these policy constraints.
- The DIMS team is operating under an NDA with the Ops-Trust organization for access to the source code for their portal. In 2014, the Ops-Trust developers released the source for the initial portal on GitHub (https://github.com/ops-trust/portal.git). Other information not made public yet cannot be released without their permission.

Note: Farsight Security has been working on a reimplemntation of the original Ops-Trust portal system, known as Trident, and plans to release it in open source form in 2016.

- The DIMS team is operating under export control restrictions that apply to any/all encryption software used in the system. Based on consultation with UW Export Control authorities, the DIMS team will design the system such that it can be released as open source without encryption software included (but will list its pre-requisite status, where it can be obtained, and how it can be installed by the end user), or will deliver pre-installed/configured versions of the system only under export control restricting agreements negotiated by the appropriate authorities at the UW.

3.3 Description of current system or situation

There are gaps in functionality in the existing sub-systems that DIMS is intended to address. The three primary sub-systems are: (1) the current PRISEM system; (2) The CIF database; and (3) the Ops-Trust portal; Each of these will be examined in turn.

3.3.1 The PRISEM System

- Event collection, correlation, archiving
- Distillation of hundreds of alerts per day from (low) tens of millions of events per day
- Integrates the NetFlow *Botnets System* behavioral detection capability
- Requires intensive administration and coding when provisioning new tenants
- Proprietary vendor portal the principal user interface

The PRISEM system works by collecting logs from each participating site, and in some cases also processing NetFlow V5 records with the *Botnets System*. At its most basic, the data flow for any given PRISEM participant site from participant to central collection and processing initially worked as shown in Figure *Original PRISEM architecture* (source: presentation on PRISEM circa 2012).

Internally, the event data collection flow at a single site looks something like Figure *Syslog Event Collection*.

PRISEM is the first regional government collaboration in the United States to enter into a Cooperative Research and Development Agreement (CRADA) with US-CERT to receive de-classified IOCs. The intent is to receive and send these indicators using MITRE Corporation's Structured Threat Information eXpression (STIX) format. The goal is to eventually link the IOCs with Tools/Tactics/Procedures (TTPs) and Courses of Action (CoA) to provide actionable intelligence to PRISEM members (see Figure *Relationship between STIX Elements* – original source: Bret Jordan, Blue Coat Systems).

The PRISEM system has demonstrated that sharing event logs within a trust community improves the situational awareness across regional SLTT government entities, that collaborative response improves everyone's capacity to

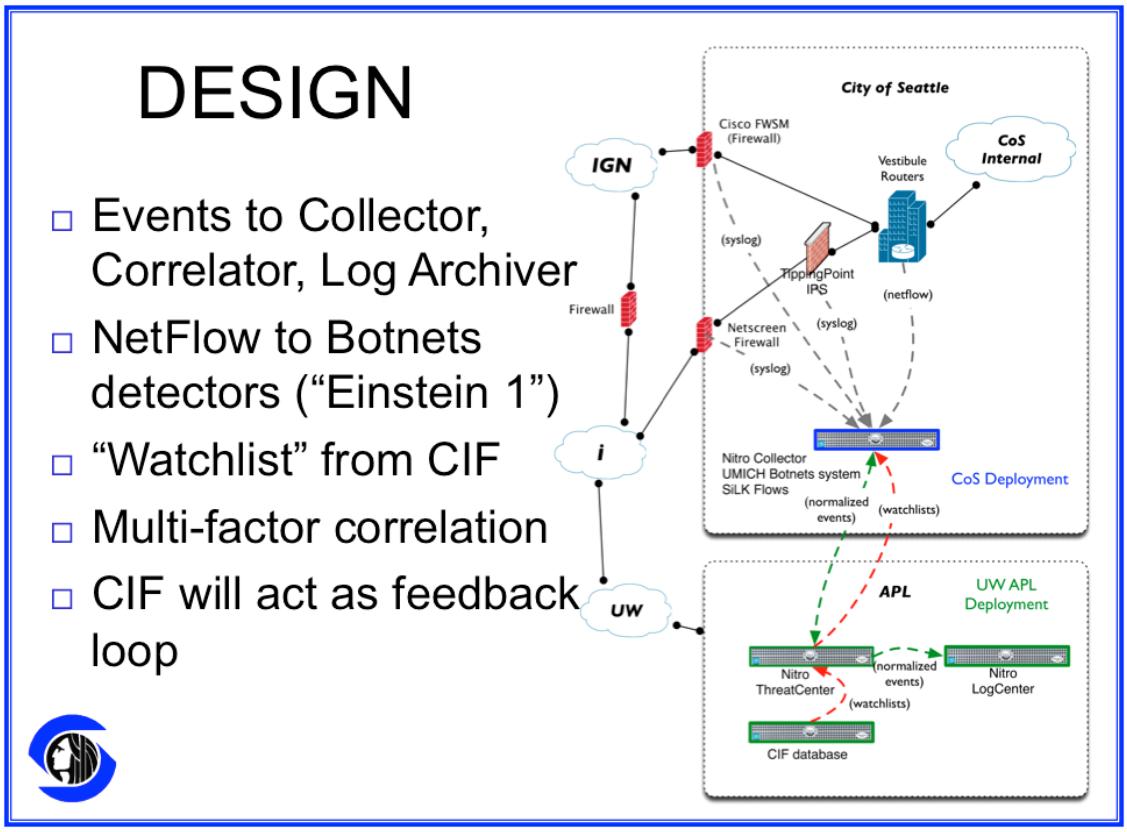

Fig. 3.2: Original PRISEM architecture

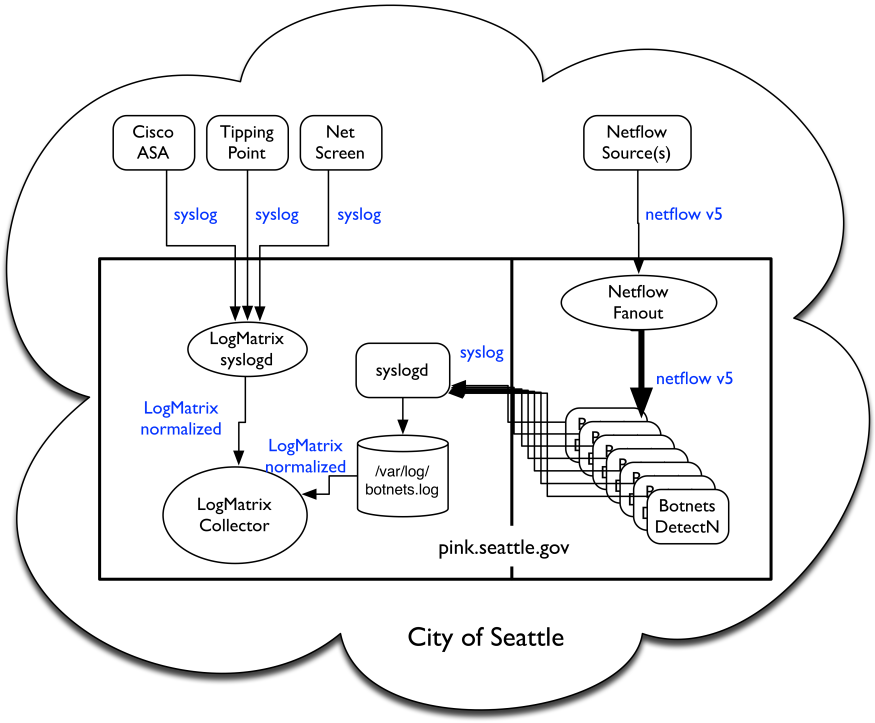

Fig. 3.3: Syslog Event Collection

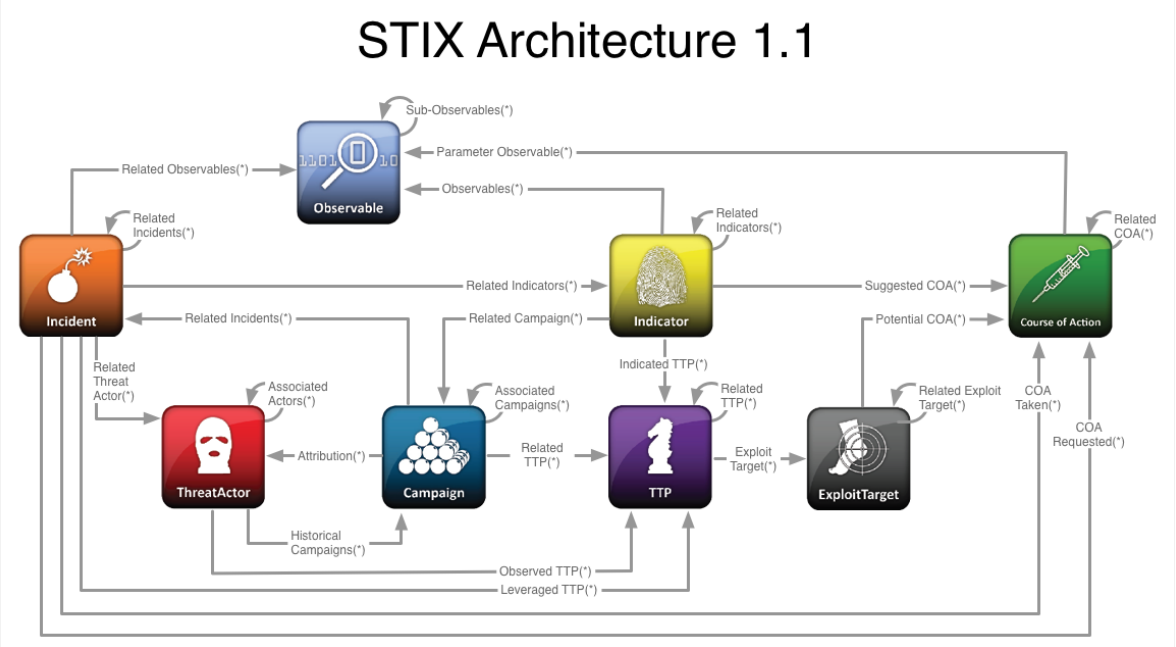

Fig. 3.4: Relationship between STIX Elements

respond and recover, and that situational awareness reports being fed back to the federal government through participation in Fusion Center activities. There are as many as five regional SLTT collaboration efforts that the PRISEM leadership has interacted with and who have expressed an interest in replicating what has been done within PRISEM (see Section *Users/Affected Personnel for New System*).

There are limitations in what PRISEM is capable of doing, primarily based on the commercial off the shelf SIEM system at its core, and the reliance on a proprietary vendor portal for the user interface that PRISEM participants use on a daily basis. There is no flexible and secure real-time communication vehicle that PRISEM participants use on a regular basis, and interaction among PRISEM participants and analyst resources could be much higher. Also related to the use of the vendor portal is a limitation on the visualization and analytic capabilities. The portal only supports what the vendor has programmed it to support. There is no easy way to integrate newly developed features, visualization tools, or analytic algorithms that operate on the PRISEM datasets.

3.3.2 Collective Intelligence Framework (CIF) Database

- "Indicators of Compromise"
- Hashes of malicious software
- IP addresses, *CIDR* network address blocks, and DNS domain names associated with malicious activity (e.g., from sandboxes)
- Builds context about attacker activity over time
- Produces feeds of indicators for watchlists, searching hard drives, rules for security devices, etc.

CIF provides a database of historic IOCs obtained from feeds that it consumes on a regular basis. In turn, CIF produces feeds of IOCs that can be used for watchlists, access control lists, IPS rules, etc. The PRISEM system uses CIF to produce watchlists that are used by the Python based *Botnets System* detectors processing real-time NetFlow V5 records sent from network devices for real-time detection of suspect flows. CIF correlates data in its tables, associating IOCs from multiple sources, as well as enriching the data by looking up ASNs, domain name to IP address associations, etc. Users can enter IOC data using CIF's browser plug-in, the CIFglue application from Verizon, or through the CIF API.

The PRISEM system also processes "SEARCH" records that are added to CIF when someone searches, putting those IP addresses or CIDR blocks that are searched for, but produce no results, into a watchlist. A more accurate way to do this is to have users explicitly put suspicious IP addresses or CIDR blocks into CIF with special tagging that is then used to generate a watchlist.

While not a lack of features in CIF, per se, the way CIF is being used is lacking in potential. While the PRISEM uses CIF to generate watchlists for real-time network flow detectors, and creates a special watchlist for "SEARCH" records as described above to watch for highly suspicious events, PRISEM users (and the vendor portal) are not taking advantage of the full power of watchlists because users must know how to manually enter data using one of the secondary CIF-specific mechanisms listed above as the vendor portal does not currently provide this ability.

CIF is also not being used to store security event information related to alerts that are positively identified by analysts as being true-positive indicators of compromise (or confirmation of IOCs sent to the system or entered manually by analysts.) Were these events to be stored, they would be correlated with other IOCs and could be published as a feed to interested outside parties.

3.3.3 Ops-Trust portal Code Base

- Handles adding users by nomination + vouching workflow processing
- Segregates trust groups (public or hidden) per administrator defined policy
- Facilitates encrypted communication via email, and out-of-band contact via phone, IM, etc.

- Provides a secure wiki for holding information contributed by users and other group knowledge
- Holds attributes about users:
  - Name, nick-name (handle) to identify them in wiki
  - Telephone number for out-of-band communication
  - Closest airport to facilitate meeting in person when on the road
  - PGP (or GPG) encryption key
  - Instant messaging system username

The Ops-Trust portal currently does a good job of the nomination and vouching workflow that allows user accounts to be set up and attributes populated. It then does a good job of segregating trust groups from each other, including facilitating encrypted email communications and storing data in a wiki.

There are several limitations to the way the Ops-Trust portal works and is used. All IOC data is passed around at present is in arbitrary forms (ASCII text columnar data in random field orderings, CSV files, PDF files, etc.) and may be in the body of an email, as a MIME attachment, or in a file specified by a URL in the body of the message. Often long lines of columnar data get wrapped and are difficult to read or parse with scripts. Cutting/pasting into security systems is difficult, if not impossible when thousands of lines of data are included in some random field in a large columnar list. Traffic Light Protocol (TLP) tagging is done in random ways (if done at all), and TLP tags in the body of a message do not get included when an attached file is saved to disk. The subject line of emails includes the list and it, and the list trailer at the bottom of the email, must be manually scrubbed when forwarding a message off-list. Users must read every message in a thread in order to keep up on new data that may involve hosts or networks that the reader is responsible for protecting, and widespread and rapidly progressing events can generate dozens or even hundreds of messages in a day, which is difficult to keep up with.

3.4 Users/Involved personnel for Current System

The current PRISEM system has the following sets of users and involved personnel:

- Participating sites are mostly contributors of event log data, and consumers of alerts and reports. They receive notification from either a managed security service vendor's Security Operations Center (SOC) staff, or from the primary analyst working out of the Seattle Fusion Center.

  Select participants in the existing PRISEM system will be involved in requirement collection, test and evaluation, and will be the initial users of a DIMS deployment.
- The current PRISEM principal analyst who interacts with the Seattle Fusion Center will contribute to requirements (primarily in the form of user stories), and will assist with test and evaluation of DIMS.
- A research scientist at the University of Washington (also the PI on this contract), who helped design and test capabilities in the original PRISEM system, will contribute technical architectural design, requirement definition, test and evaluation, documentation, and initial user training on the DIMS system.

3.5 Support concept

The current PRISEM system has been supported through grant funding, support for hosting hardware by entities at the University of Washington, and contracting with a commercial managed security service vendor with working experience with the underlying commercial SIEM system originally chosen for use by PRISEM. This system is known as *Log Matrix* and is an end-of-life product now owned by Intel subsequent to their acquisition of McAfee.

CHAPTER 4

Justification for and nature of changes

4.1 Justification for change

Knowledge is becoming a critical success factor for organizational performance. Many public and private organizations are sharing knowledge as one of the means to collaborate and gain sustainable competitive advantage over these threats. Advances made in information and communication technology (ICT) is aiding these efforts. The need for infrastructure protection and real-time to near-real-time automated response to cyber threats to enable expedient top-level decisions has become imperative. However, a widely accepted framework for visualization, analytics, situational awareness, enabling intraregional response to shared threats does not exist today.

To address these concerns, a system called the Distributed Incident Management System (DIMS) will be built. DIMS will be based mostly on existing technology, much of it from the open source software development community, and leveraging emerging standards. The primary users of DIMS are the Computer Security Incident Response Teams (CSIRTs) who need to maintain the security and functionality of a diverse and complicated, yet institutionally critical cyber infrastructure. DIMS will be based on open source technology and standards.

4.2 Description of needed changes

As mentioned in the previous section, MITRE has been working with US-CERT to develop standards that enable the kind of response and recovery process called for by EO 13636 and PPD 21. To that end, they have illustrated how STIX can be applied to four specific use cases that bridge local to national response. These use cases (shown in Figure *STIX uses cases (from MITRE)*, taken from the STIX web site) are: *Analyzing Cyber Threats* (UC1); *Specifying Indicator Patterns for Cyber Threats* (UC2); *Managing Cyber Threat Response Activities* (UC3); and *Sharing Cyber Threat Information* (UC4). *[The12]*

MITRE defines *observable* as, "[an] event or stateful property that is observed or may be observed in the operational cyber domain, such as a registry key value, an IP address, deletion of a file, or the receipt of an http GET. STIX uses Cyber Observable eXpression (CybOX) to represent Observables." The PRISEM system collects logs that contain the IP addresses of the source and destination of events and flows, along with other information about specific security events (sometimes including domain names, URLs, services being used, and observed attack signatures).

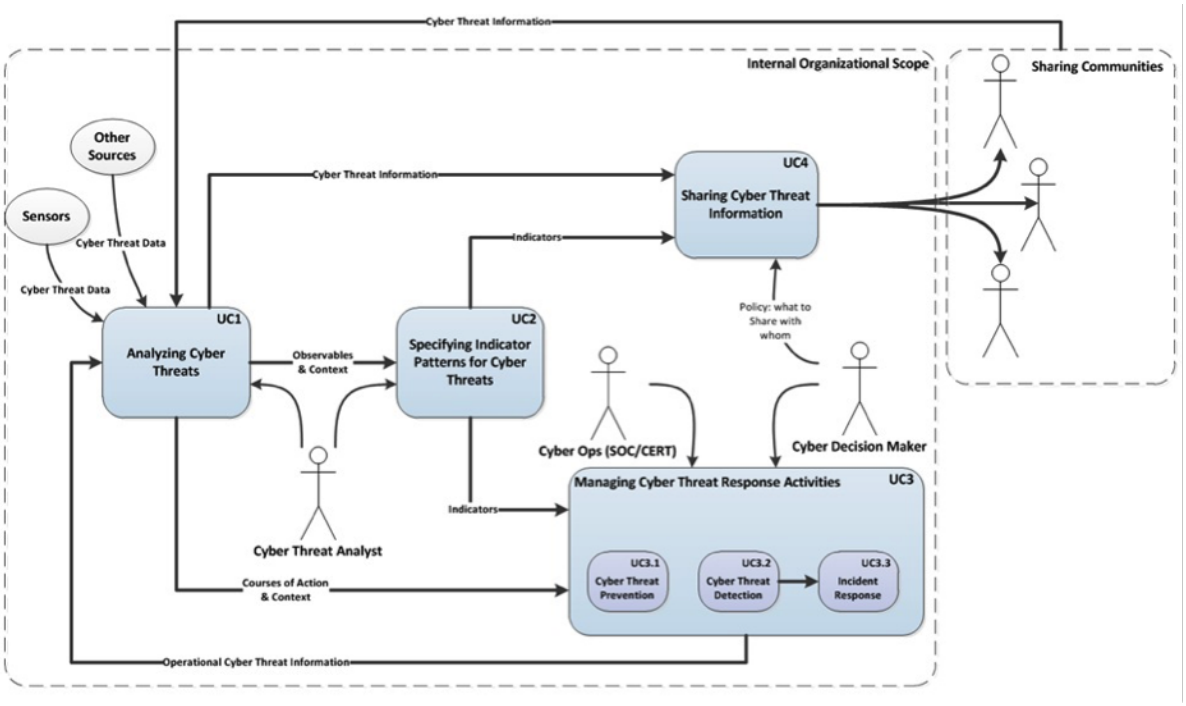

Fig. 4.1: STIX uses cases (from MITRE)

MITRE defines *indicator* as, "[a] pattern of relevant observable adversary activity in the operational cyber domain along with contextual information regarding its interpretation (e.g., this domain has been compromised, this email is spoofed, this [*cryptographic hash* of a file] is associated with this trojan, etc.), handling, etc. An Observable pattern captures what may be seen; the Indicator enumerates why this is Observable pattern is of interest." (STIX FAQ #B1) One job of an analyst using the PRISEM system is to take *indicators* that are shared by outside sources, which are used to trigger alerts within the PRISEM system, and connect them with those logs that include related observables and other context (such as the information stored in the Collective Intelligence Framework database) and distill them into analytic products like situational Indicators of Compromise, or IOCs, can also be described as "a forensic artifact or remnant of an intrusion that can be identified on a host or network. [IOCs] tie to observables and observables tie to measurable events or stateful properties which can represent anything from the creation of a registry key on a host (measurable event) to the presence of a mutex (stateful property)." *[Gra12]* IOCs can include several pieces of raw intelligence that manifest at various points in time on information systems under attack, including "MD5 [and other *cryptographic hash* values for files], File names, Packer types, Registry keys, Mutexes, DNS strings, and IP Addresses." *[Man11]*

IOCs are the lowest-level pieces of evidence used to paint a much larger picture as part of the response and remediation process. *[Ald12]* They are the needles to attempt to find in a haystack, not a request to go find needles. Many of these indicators are found within the file system of a compromised computer, while others can be found in network flows and server logs that include transport and network layer information (e.g., IP addresses and IP protocol and port numbers.)

A workflow or workflow process is the set of steps that someone goes through to perform a complex task, such as fulfilling an order for an online purchase, or performing forensic analysis of event logs and network flow data to confirm compromise, determine root cause, and learn the extent of a breach. Microsoft describes it this way: "Workflow is fundamentally about the organization of work. It is a set of activities that coordinate people and/or software. Communicating this organization to humans and automated processes is the value-add that workflow provides to our solutions. Workflows are fractal. This means a workflow may consist of other workflows (each of which may consist of aggregated services). The workflow model encourages reuse and agility, leading to more flexible business processes." *[Mic]*

In the case of the forensic analysis process that underlies response as described above, the workflow is fractal in terms

of including other workflows, but is also a recursive process. This process can start with one or more IP addresses or network address blocks that are suspicious. This can lead to a set of potentially compromised computers who had communication to that single IP address. Looking at the flows to/from those suspect computers results in a larger set of potentially malicious computers that are related to the first IP address, but were not known at the start. The developing network of malicious activity grows with each iteration in the discovery process and each new search result builds on previous knowledge. As the network increases in size, the analyst wants to filter out known good hosts, and highlight the known bad hosts, in order to find new suspect hosts to evaluate (and then hopefully move to the known good or known bad sets.) Keeping track of the growing body of known good and known bad is a requirement of the workflow for this discovery process.

The objective of the DIMS system is to support the following high-level missions and needs, which incorporate the four use cases described above as defined by MITRE:

1. To facilitate collaborative response to shared threats by supporting real-time and near real-time communications, situational awareness in graphical and text report formats, and role-based controlled access to security event and alert data housed in a shared SIEM system. (UC1 and UC3)
2. To provide a framework for visualization and analytic tools that result in a shared view of common threats, in a manner that compares and contrasts each participant with others in the system to help them understand whether certain threats are widespread and common, or may be targeted to a specific sector, organization, or physical locality. (UC3)
3. To facilitate the real-time and near real-time operational sharing of actionable information in the form of structured IOCs and Observables that support triage, response and recovery, and determinations of events of such criticality that they require reporting to federal authorities. These IOCs and observables may come from US-CERT (as part of the CRADA between US-CERT and the PRISEM project), may come from other trust groups (be they sector-specific, regional, or self-organized), or may come from federal law enforcement agents in the local field office. As IOCs and Observables are linked with TTPs and COAs (see Figure *Linking minimal subset of STIX elements from Observables to COA* for an example of the minimial linkages necessary to operationalize IOC and Observable sharing), the users can more quickly and efficiently respond and recover. (UC2, UC3, and UC4)
4. To facilitate tracking of remediation efforts across participants. It is a common occurrence to receive a report with a list of IP addresses and/or domain names of suspected compromised or abused hosts. Having a mechanism to automatically determine which IP addresses are of interest to which participants by comparing those addresses to assigned network blocks or top level domains makes it easier to know when attention should be paid to data coming in to the system. Similarly, after remediation it is possible to toggle the status of these hosts and automatically keep track of when a site has completed cleanup, what percentage of known compromised hosts have yet to be mitigated, and how quickly they are being cleaned up. This information speeds up overall response and provides metrics by which to compare process improvements over time. (UC1 and UC3)
5. While not directly mapping to one of MITRE's use cases, the DIMS effort is intended to enable integration of complementary open source security tools and put these tools back into the community as open source tools, and/or transition these tools into commercially available products that advance the state of the art in distributed incident response.

4.2.1 Ops-Trust portal Code Base

The principle mechanism lacking from the Ops-Trust portal is the ability to pre-process IOC data sent by users so as to notify each user when a thread pertains to them (because IOCs match pre-defined lists that the user cares about), and more specifically, which email messages contain IOCs of interest. The data necessary to do such filtering and altering is not stored in the Ops-Trust portal database, nor is there a standardized mechanism for passing machine-parseable data into the portal to facilitate workflow automation. The Ops-Trust portal is also monolithic and focused on managing the trust groups and users, not on making data analytics and visualization capabilities available to help process the IOC data that is available throughout the user base. It does not have capabilities to anonymize data, nor to

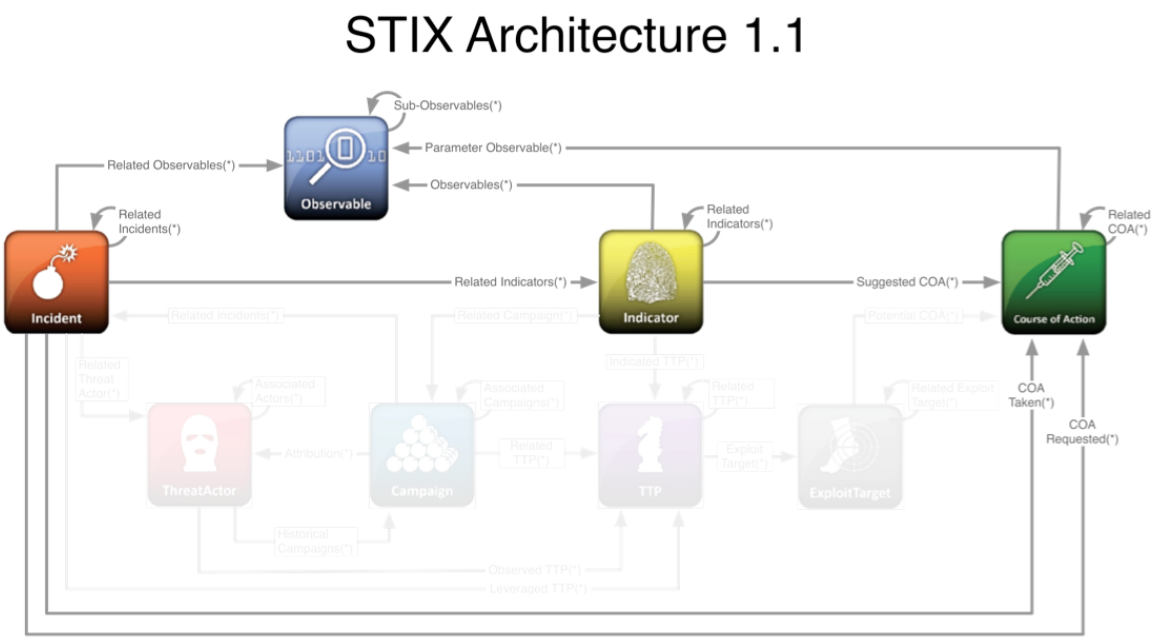

Fig. 4.2: Linking minimal subset of STIX elements from Observables to COA

associated TLP tags with data such that filtering and anonymization does not rely solely on humans knowing when/how to filter and anonymize data, and on them never making mistakes.

The Ops-Trust portal, written in Perl with a PostgreSQL database backend, needs to be refactored, using a model-view-controller framework (MVC) framework such as Catalyst (http://www.catalystframework.org/), to separate the front end UI capabilities from the back-end database and portal workflow processes so as to provide an API that alternate UI components can access via a standardized mechanism such as a RESTful HTTPS interface. The UI needs to be refactored to improve usability and provide access to both user and administrator functions. It needs to have additional user attributes added to facilitate the filtering and notification process described above, as well as to have workflow processing features added to perform some of the manual filtering and searching capabilities. The account management features need to be extended to support AAA and RBAC features that use mechanisms such as roles and TLP tagging to ensure exported data is filtered and/or anonymized in accordance with user-defined policies. Once the MVC conversion has been completed, and some of the additional attributes and features necessary to semi-automate information sharing, an application penetration test needs to be performed to satisfy requirements of the authors for publicly releasing the code as an open source project.

Adding features to enable trusted sharing of machine-parseable IOCs between instances of the Ops-Trust portal makes it possible to scale trusted information sharing to a larger population than the existing Ops-Trust group is capable of growing. Having additional attributes for users enables workflow automation of notification of IOCs relevant to their constituencies, which speeds response. Eventually, features that ensure the chain-of-custody and provenance of security data that can be used as evidence in criminal or civil legal proceedings, combined with the machine-parseable nature of the data exchange, will facilitate reporting computer crimes to law enforcement in a manner that speeds their investigations and helps more accurately scope and prioritize investigations.

4.2.2 Collective Intelligence Framework (CIF) Database

It is unknown how much data can be put into CIF before it reaches performance or storage limits. As part of the PRISEM deployment of CIF, mechanisms were put in place to regularly log the sizes of certain database tables and the

database itself, and to log the amount of time it takes to pull feeds from outside sources, to perform correlation, and to index database tables (all processes that run from *cron* on a scheduled basis). This information has only been used to answer questions at given points in time, but the intention was to perform linear regression on this data on a regular basis to estimate when resource limitations will be hit (e.g., when the disk drive is expected to be filled to 100%, or when the CPU processing capacity approaches 100% on a continual basis.) This would allow better monitoring of resources, tuning of system parameters, and estimation of hardware capacity required as the PRISEM population increases. All of these features would be made available to the CIF developers to extend the capability of all CIF users to be pro-active about their deployment infrastructure.

As CIF is a "work in progress" and constantly undergoing development, the community of users is often called upon to help identify bug fixes and feature additions that can be made available to the CIF development team via *Git* "pull" requests. This helps improve the generally available release of CIF and minimizes the need to maintain add-on patches independent of CIF releases. Since the intention of DIMS is to be replicated in many regions, each of which constitutes a different mix of participants, security data sources feeding the central SIEM, etc., mechanisms to better identify capacity requirements and monitor runtime resource usage for minimum downtown becomes critical. The same machine learning algorithms used for resource monitoring are also useful for clustering and classification of security event data, so their implementation in a generalized framework increases the flexibility of their application.

4.2.3 The PRISEM System

The underlying inter-process communication added to the PRISEM system in recent months provides a flexible and extensible mechanisms for Remote Procedure Call (RPC) invocation, as well as logging of information about queries and response times that can serve to estimate wait times for longer queries. This message bus architecture is also programming language agnostic, operating system agnostic, and is using a structured command structure that allows self-description of the data being sent between programs to facilitate merging results from multiple processes (e.g., the "identify friend or foe" capability, anonymization and statistics, partitioning and filtering based on participant network allocation attributes, etc.) A new user interface that supports all of these capabilities in a flexible framework architecture will allow seamless integration between any SIEM product, any vendor portal, and any open source security tools that are appropriate for processing the kind of data held within PRISEM.

Adding a layer of abstraction above the SIEM and vendor portal allows flexibility for any SIEM, or any managed security service vendor, to be employed to build a PRISEM-like regional collaborative group. There are many competitors in this field, and none of them combines the features of universal compatibility, affordability across the full range of small to large SLTT collaborative groups, and ease of migration or interoperability as regional collaborative groups spontaneously form and grow. What do you do if two groups using two different SIEM products and two different vendor portals wish to merge? What do you do if the SIEM you are using reaches its end-of-life and is now longer supported, necessitating a migration of over a year's worth of normalized log data to be translated to a new product? What do you do if a group decides they want to replicate the PRISEM model, and now has to scope out a SIEM deployment and/or managed security service vendor contract for provisioning and support? These are all realistic questions, very hard to answer in the short term, very costly to enter in to, and take a significant effort to reach a go/no-go decision point. An abstraction layer that focuses on standardized data interchange, vendor-agnostic interfaces to data, and an open framework for new features, solves many of these problems and provides the affordability, flexibility, and scalability that is needed to reach national scope.

4.2.4 Summary of the capabilities gap

The principal high-level gaps that exist in supporting the missions described in the previous section have to do with the availability and affordability of tools that support those missions. Each of these tools have limitations or impediments to their use:

- There are managed security services that could be engaged to handle all security incident response and forensics. The cost of these services is prohibitive for all but the most serious incidents with potential losses that rise to the level of existential threats to the viability of the enterprise. The availability of affordable open source tools to improve response and recovery is a gap that DIMS is intended to fill.

- There are agent-based systems and network-based that can provide the level of detail and pervasive collection of event data at the host, server, and network levels. These, too, are prohibitively expensive. They only work in environments where policy can dictate the deployment of agents on all end hosts and servers, and where network topology and administrative responsibility at the enterprise level is such that one group can deploy, manage, and interact on a daily basis with the security system. Most SLTT government sites cannot afford to have this level of in-house security monitoring and response capacity. At present, even if one site in a region can afford such capabilities, their use is limited to protection of that site alone and there is little benefit to other inter-related entities in the region (hence the need to share not only IOCs and Observables, but also Course of Action and analytic results.)
- Most SIEM systems focus on the problem of collecting and correlating millions of events per day, distilling them down to a reasonable (N<=100/day) level, and directing them to the entities with administrative control over the system identified in the alerts. Correlation across a confederated population is not typically done (most deployments are for one enterprise, perhaps with multiple business units under the same top level corporate structure). These systems are also primarily focused on detection and alerting on input of events, not on after-the-fact triage and respond/recover operations. When they do support forensic analysis of past events, these systems typically do not support confederated cross-organizational correlation and collaborative response (e.g., by sharing analysis between multiple enterprises, or distributing Course of Action information.)
- The existence of the Ops-Trust community proves that volunteers can self-assemble to respond and react to issues that impact everyone on the internet, but these groups frequently operate on email and chat communication channels that are unstructured, ad-hoc, and are very difficult to keep up with. Unless one reads every message in every email thread, extracts all attached files or processes all in-line data, and manually searches for IOCs and Observables that can be manually used to search data sources that that person controls, the benefit of information sharing is lost. And for any emergent situation of global significance, the threads are many and the messages in each thread can flow for days or weeks. It is impossible to keep up with this without moving to structured data and machine processing to identify messages of interest.
- There have been many formats for structured security data sharing developed over the years. Each one has seen a similar lifecycle, where there is interest and excitement at the start of the project, a slow deliberative process of developing the standard, going through the process of vetting and acceptance of the standard by an official body, and then a push to get the industry and researchers to adopt the standard. STIX may encounter this same fate. It is too early to tell. What some (like Wes Young, developer of the Collective Intelligence Framework) suggest as an alternative is to "blow up the standards process" and simply implement something quickly, get it used by as many people as possible, adapt and modify it to address limitations that are encountered, and keep moving forward. "We believe traditional standards processes not only have a high barrier to entry, but are often slow and use the design by committee approach. We believe the best way to create a protocol is from the ground up using CONOPs. Push design out to the edge and let operations influence design in real-time." (CSIRT Gadgets Foundation web site)

4.3 Assumptions and constraints

The following assumptions and contraints are applicable to the changes identified in this section:

- The use of open source tools brings with it the challenge of integrating a number of code bases that are written in different programming languages, have different coding styles, differing interfaces and input/output data formats and mechanisms, run on different operating systems, have specific and possibly incompatible pre-requisites, may have duplication in (or conflicting choices of) database mechanisms, and may have little or poor documentation.
- Attempting to balance all of the differences mentioned in the previous bullet will push all team members to the limits of their technical abilities.
- Hardware, network resources, and data center limitations can cause friction due to limitations on access to data center facilities, the distributed nature of the development team, and where certain services can/should run.

CHAPTER 5

Concept for a new or modified system

5.1 Background, objectives, and scope

One of the objectives of DIMS is to combine the best features of several open source projects using a *framework* model that integrates these components into a coherent whole. All of these systems were built by groups independently of each other, often with volunteer effort, or with limited budgets within corporations that chose to make these tools available as open source to encourage use by the security community.

One of the primary challenges faced by the DIMS team will be to move beyond the mindset of installing and configuring a small set of discrete open source packages on a single workstation and using the tools like a normal security operator. This mindset is limited in that it assumes stasis, or at least little change or modification beyond that provided by regular patches or releases from the open source author.

Producing a *framework* means using automated build processes, commonly known today as *DevOps* (see What is DevOps?) as a method of automating the build+configure tasks faced by system administrators, and using Continuous Integration as a method of managing the source code for programs and system configuration, pushing those changes and compiled programs into running systems.

As much as possible, DIMS will be built through the (re)use of open source components used by other projects that are being integrated into the DIMS framework. For example, the Collective Intelligence Framework (CIF) v2 and the Mozilla Defense Platform (MozDef) both employ the ELK stack and RabbitMQ in their demonstration implementations, and the original PRISEM distributed data processing tools also used RabbitMQ. Rather than have two separate instances of Elasticsearch running in virtual machines or containers for MozDef and CIF, and two separate instances of RabbitMQ in virtual machines or containers for PRISEM tools and MozDef, a common Elasticsearch cluster and RabbitMQ cluster would be set up and shared with these and any other open source tools that someone would want to add in later. (Another example of a system made up of multiple components, packaged together into a single easy-to-install package, is the GRR Rapid Response system.)

Figure *Recombination of open source systems* illustrates the thinking behind this DevOps/CI mindset, and how it can be applied to build DIMS. The upper half of the figure represents (conceptually, not in precise technical terms) the way that open source systems are commonly bundled together. From left to right are the Collective Intelligence Framework described in Section *Collective Intelligence Framework (CIF) Database*, MozDef, some of the PRISEM system components described in Section *The PRISEM System*, and the ops-trust portal described in Section *Ops-Trust portal Code Base*. From top to bottom in this conceptual model are the common components of application user

interface (in this case, a RESTful HTTP/HTTPS interface), a message bus mechanism for inter-process communication that can span computer systems, a database storage mechanism, and a base operating system within which all of these components are installed.

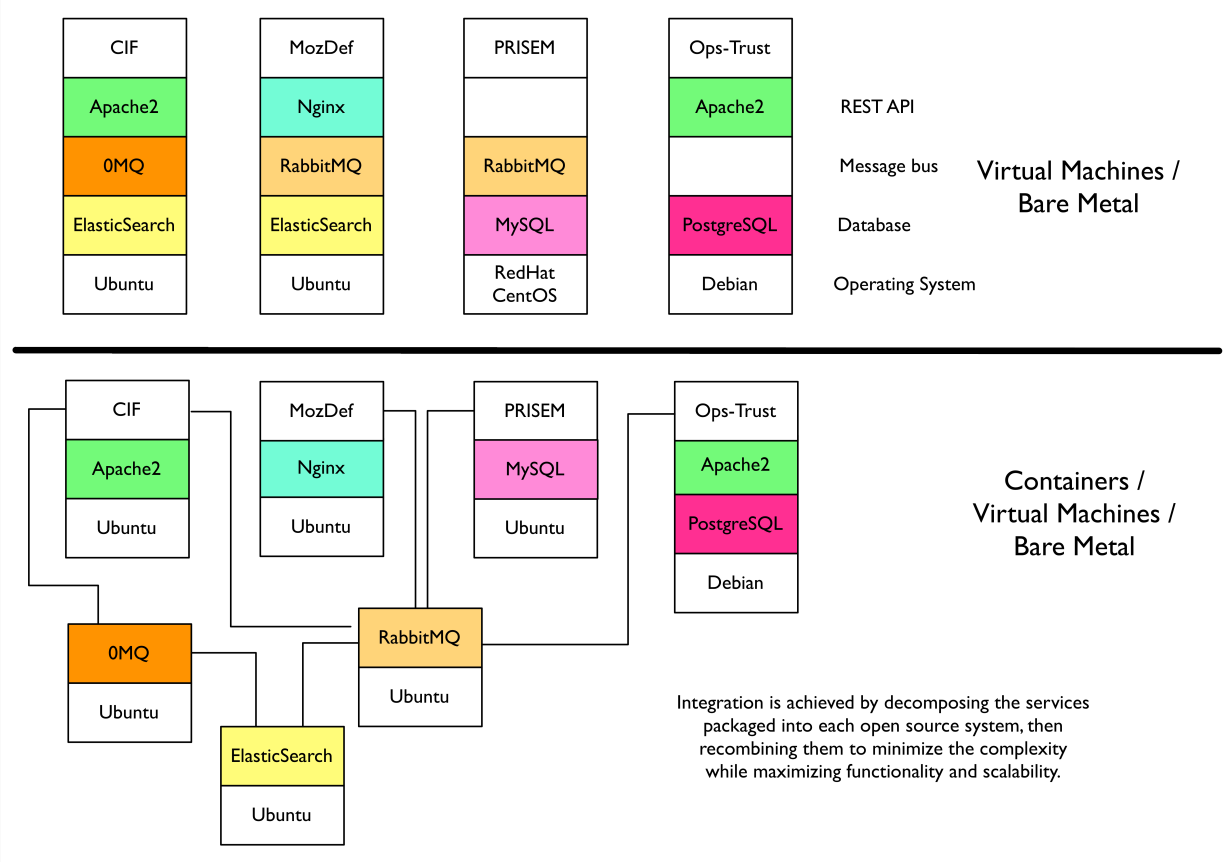

Fig. 5.1: Recombination of open source systems

The bottom of the image depicts, again conceptually, how you would rip apart or docompose the subsystems in these *packaged* deployments, and turn them into discrete component services that are contained in smaller units. By compartmentalizing services in this way, it may be easier to integrate several open source packages that may have conflicting requirements for base operating system type, operating system version, libraries (and their versions), or configuration and tuning parameters for shared services (like the PostgreSQL database). In the bottom of Figure *Recombination of open source systems* one Elasticsearch cluster, and one RabbitMQ cluster, can be implemented and shared by multiple components (rather than having two seperate small clusters in two separate virtual machines or bare-metal machines. This would allow linear expansion of these clustered services as needed for growth. (It could even be possible to elminate one of the two message bus systems, either RabbitMQ or ZeroMQ, to further simply the architecture.)

5.2 Operational policies and constraints

5.3 Description of the new or modified system

Figure *Overview of DIMS System* depicts a high-level diagram of the system architecture for the DIMS system. DIMS provides a user interface layer on the front end, as well as a data processing layer on the back end, that integrates with two existing systems.

The first is the Security Information Event Management (SIEM) system at the core of the PRISEM project, and the technologies associated with it to perform behavioral detection of malicious activity from network flow data and support forensic analysis of historic data to respond and recover from attacks that evade detective mechanisms. This system collects and processes security related events and network flow data and supports a collective approach to responding and recovering from security events.

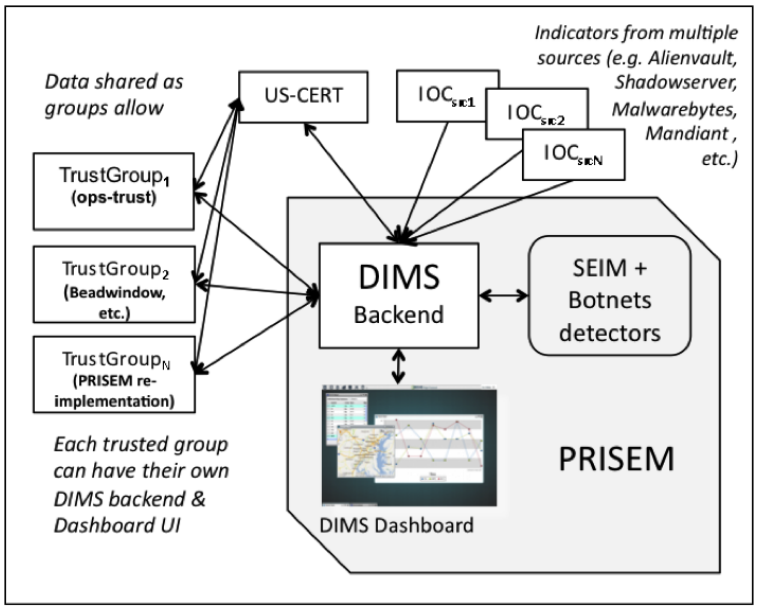

Fig. 5.2: Overview of DIMS System

The second system is the Ops-Trust portal system, used by a community of several hundred computer security professionals with operational and research roles in industry, government, and academia. This system is primarily designed to facilitate trust group maintenance and communication to deal with emerging threats and events of international scope.

The DIMS software will bring these two systems together into a collaborative environment for shared analysis and shared response of shared threats, both within a regional trust community, as well as across multiple such trust communities in other regions. Through vertical sharing of indicators of compromise from US-CERT to the regional level, and lateral sharing across regional entities, the objective is to scale actionable information sharing to state, local, territorial, and tribal (*SLTT*) government entities across the United States, and extend the sharing to international trust groups who make up the global fabric of the internet.

Figure *Data Flows Between Stakeholders* depicts the data flows between a subset of the stakeholders who will be using the DIMS software system. The solid lines depict data that has the highest degree of sensitivity and trust, often being transmitted in un-redacted form (possibly tagged with TLP indicators for most restricted sharing). The dashed lines depict data flows that are at lower levels of trust, and may be transmitted only in redacted form (possibly tagged with TLP indicators for the least restricted sharing). The type of data shared may be structured IOC and Observables in STIX format, Course of Action information in either PDF or structured format, *Situational Awareness Report* (SITREP) documents that describe observed campaign level activity at a high level, possibly with structure data containing IOCs or Observables to assist recipients in searching for related activity, and incident reports that may similarly be a combination of human-readable PDF and machine-readable IOCs/Observables. There are two types of data that will be shared in most use cases: high-frequency, high-volume, automated data feeds of *reputation* data and IOCs/Observables coming from analytic and research groups; low-frequency, low-volume, manually triggered bundles of IOCs/Observables, Course of Action information, and/or high-level SITREPs for specific incident-level up to campaign-level activity. The DIMS software, layered on top of the Ops-Trust portal system, will facilitate production of these reports and transmission/reception of structure data files and facilitate automated processing of the structure data files to pre-process data for an analyst to consume when ready, rather than forcing the analyst to do a lot

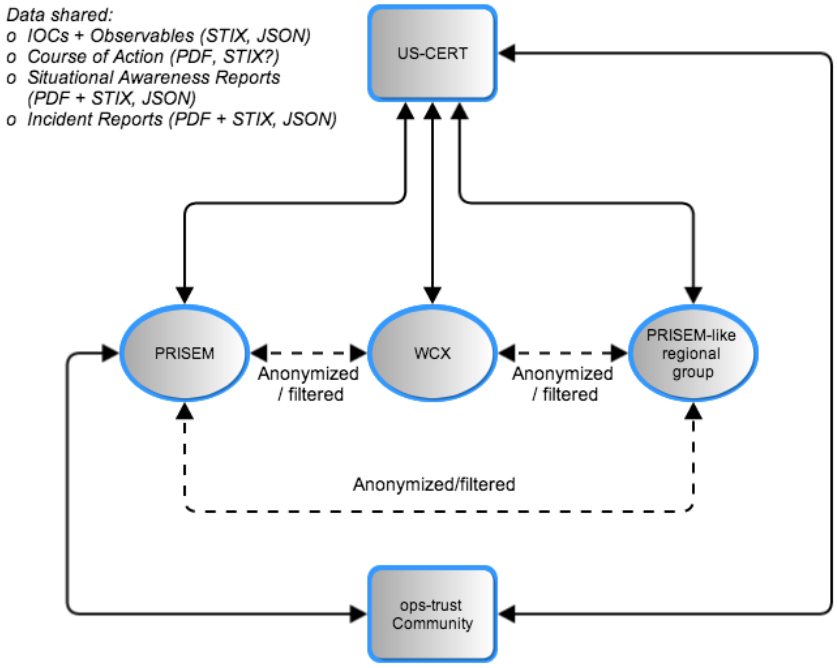

Fig. 5.3: Data Flows Between Stakeholders

of work manipulating files, processing their contents, and manually entering data into report generation front ends in web based portals.

5.4 Users/Affected Personnel for New System

The full list of stakeholders and prospective users of the new system includes:

1. *PRISEM participants*: Existing participants in the PRISEM project in the Puget Sound will be the primary users of the DIMS system. DIMS is being designed to provide them with advanced mechanisms for rapid response, situational awareness, and communication within the trusted group. Next highest priority is to provide structured data interchange between the existing Ops-Trust portal and the DIMS system, allowing lateral sharing of IOCs and observables between the existing Ops-Trust community members and PRISEM participants as allowed by policy (or with redaction and/or anonymization, as appropriate.) Some features added to the Ops-Trust portal by the DIMS project team will be integrated in such a manner that they are available to Ops-Trust members without having to use the DIMS front end software. Those users who are not part of the existing Ops-Trust community, or Ops-Trust members willing to learn a new interface, can use the DIMS front end and will have access to a larger set of features than are available via the normal Ops-Trust services.
2. *PRISEM Administrators and DIMS developers*: Related to the PRISEM membership is an entity being formed to administer the PRISEM model in the form of a not-for-profit organization responsible for daily operations, system administration, provisioning of SIEM collectors and SIEM configuration, training, etc. This entity is still being formulated and does not exist today (however it is likely to exist before the end of the option year for the DIMS project.) The DIMS developers will also serve as system administrators, trainers, and user support for the initial DIMS deployment while the PRISEM stand-alone entity is being stood up.
3. *US-CERT*: Provides IOCs in STIX format to PRISEM participants as part of an existing Cooperative Research and Development Agreement (CRADA) between US-CERT and the PRISEM project.

4. *Ops-Trust*: This is a community of several hundred operational security professionals from the private sector, academia, etc. They currently share information in ad-hoc ways, primarily through email communications and IRC chat.
5. *NCFTA*: This is a federal government and industry collaborative organization primarily focused on computer crime related information sharing and analysis. They are located in Pittsburgh, Pennsylvania, but interact with corporate and government entities from a number of countries. NCFTA has complementary needs to those of the PRISEM participant base (though focused more on investigation than day-to-day monitoring). They are eager to take advantage of features provided by DIMS that support the investigator and analyst use cases. They have offered to compare requirements and use cases to their own needs, to help test new Ops-Trust and DIMS features, and provide feedback for test and evaluation of DIMS products.
6. *Western Cyber Exchange* (WCX): WCX is a non-profit entity located in Colorado Springs, Colorado, that integrates horizontally on a cross-sector and regional basis to allow for non-traditional information sharing between government and industry. They have expressed an interest in replicating the PRISEM model and in participating in DIMS software development and testing. Web site: wcyberx.org
7. *True Digital Security*: True Digital provides network security assessments, vulnerability analysis, network security monitoring. They operate in the Tulsa, Oklahoma region. Like WCX, they have expressed an interest in replicating the PRISEM model and in participating in DIMS software development and testing. Web site: truedigitalsecurity.com
8. *United States Secret Service*: Federal law enforcement agency who would consume cybercriminal case information from victimized SLTT entities (such as the PRISEM user base an other similar stakeholder groups). They operate on a similar model to the UC1 and UC3 entities shown in Figure *STIX uses cases (from MITRE)*, only focused on criminal investigative and national security situational awareness tasks and not security operations tasks like other federated groups like ISACs.

5.5 Support concept

Efforts are underway to create a non-profit, tax-exempt non-governmental organization who is capable of engaging with SLTT government entities via inter-local agreements. This entity will operate on a self-sustaining, fee-based model that has been described by Parker Montgomery in his report, "Organization Design: A Sustainable and Self-Sufficient Model for Washington State's PRISEM Partnership" (see *Referenced documents*).

The open source tools used to create DIMS, as well as the source code and development infrastructure used to create DIMS, will all be released to the public and will be deployable on modestly priced commodity hardware. This makes for an affordable solution for SLTT government groups or other organizations who wish to participate in trusted information sharing in a scalable manner. There will be some ongoing costs associated with maintaining and administering a DIMS deployment, but the goal is to provide as much documentation as possible to keep the support costs down.

For more information, see the DIMS Commercialization and Open Source Licensing Plan v 1.7.0 document.

CHAPTER 6

Operational scenarios

This section describes several operational scenarios that illustrate the role of DIMS, its interaction with users, its interface to other systems, and the states or modes identified for the system. These scenarios include events, actions, stimuli, information, interactions, etc., as applicable that form the basis for the requirements and user stories in DIMS System Requirements v 2.9.0.

A common scenario occurring regularly today involves responding to what are known commonly as *botnets*, or distributed intruder attack networks constructed of computers infected with malicious software (or *malware*). A botnet is the name given to a set of stolen computer assets that form a distributed computer attack network capable of performing many functions for a computer criminal. These functions can include any/all of the following: Distributed Denial of Service (DDoS) attacks of various types; scanning for vulnerable hosts to infect to grow the botnet; searching computers for sensitive information (e.g., email addresses, credit card or banking information, login accounts and passwords, files containing proprietary data that are to be exfiltrated; sending spam emails; etc. This is typically accomplished by first compromising a number of computers using one of several direct or indirect methods of propagation, resulting in installation of malicious software followed by outbound (or inbound) connections to achieve command and control (C&C) of the infected hosts, or *bots*.

The role of SIEM in this context is to provide *correlation* of multiple events, not just to trigger alerts based on single detected events. The Botnets system used within the PRISEM project produces reports that summarize individual discrete events, which by themselves may be *false-positives*. Even when a score is high because of multiple alerts being generated for repeated activity, the alert may be meaningless. Or someone may have entered an indicator into the database with low confidence of suspicious activity, which made its way into a *watchlist* detector that begins to trigger events when connections are seen to the watchlisted IP address. Requiring that multiple different alerts occur simultaneously (e.g., scanning, attempted SMTP connections, and suspected botnet command and control) before the events become elevated to *alerts* has the effect of increasing the probability that the host involved is truly compromised (i.e., a true-positive alert). The analyst looking at alerts and reports must be careful to know *what the alert means*, *how it was derived*, *what its confidence level is*, and *whether it is a valid alert or not* before passing it along, or to at least reflect a low confidence or otherwise include a caveat statement unless and until other correlating data substantiates malicious activity.

6.1 Generalized Analysis Scenario

Using PRISEM components to walk through some of the steps in responding to a suspected botnet related event helps illustrate the process:

1. The analyst may start with a message that provides indicators of compromise. Figure *Email with indicator of SSH scanning activity* shows a message reporting a suspected network involved in known SSH dictionary scanning, attempting to gain access to insecure accounts.

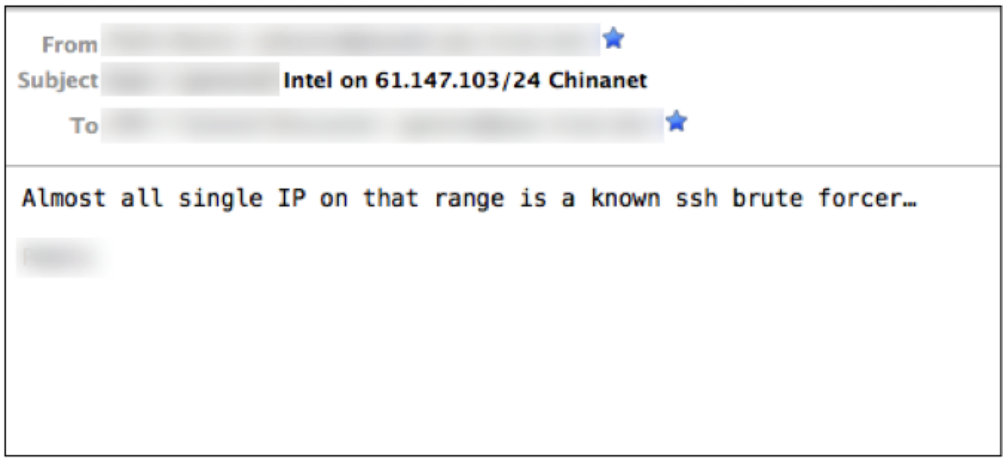

Fig. 6.1: Email with indicator of SSH scanning activity

2. The analyst can look in CIF to find what is known about this netblock. From public sources, this network block has been known for a while to be involved in SSH password-guessing attempts. (Figure *CIF lookup results for scanning CIDR block* only shows the first few fields from the CIF database.)

```
dittrich@floyd:~$ cif -q 61.147.103.0/24

Query: 61.147.103.0/24
Feed Restriction: RED
Feed Created: 2013-11-09T20:00:12Z

restriction|guid    |severity|confidence|detecttime          |address       |protocol|portlist|asn
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.149|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.165|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.166|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.167|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.168|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.169|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.170|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.171|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.186|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.187|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.188|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.189|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.190|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.191|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.192|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.208|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.209|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.210|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.211|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
AMBER      |everyone|medium  |65        |2012-07-05T00:00:00Z|61.147.103.212|        |        |23650 CHINANET-JS-AS-AP AS Number for CHINANET jiangsu province b
```

Fig. 6.2: CIF lookup results for scanning CIDR block

3. The analyst can search historic network flow records to see if there were any recent flows to/from the reported suspect CIDR block. In this case, a seven-day search does turn up some flows. The output in Figure *Confirmation of network flows related to suspect CIDR block* shows both raw output form and anonymized output using the methods described earlier:

4. The analyst may then query CIF using the web browser interface to see if this specific IP address, seen in the identified flows from the previous step, has any information about it. Figure *Output of full CIF query via*

```
[dittrich@pink ~]$ rwfind --days 7 -k -a 61.147.103.0/24
rwfind -- Sat, 09 Nov 2013 11:23:20 -0800

Found 2 flows over the last 7 days to/from the following:
61.147.103.0/24

            sIP|            dIP|sPort|dPort|pro|   packets|     bytes|                  sTime|      dur|
 61.147.103.134|  156.74.144.70| 9023|   22|  6|         1|        46|2013/11/05T04:36:05.651|    0.000|
  156.74.144.70| 61.147.103.134|    0|  778|  1|         1|        68|2013/11/05T04:36:05.653|    0.000|

rwfind -- Sat, 09 Nov 2013 11:23:20 -0800

Found 2 flows over the last 7 days to/from the following:
61.147.103.0/24

            sIP|            dIP|sPort|dPort|pro|   packets|     bytes|                  sTime|      dur|
 61.147.103.134|  [___CTYSEA__]| 9023|   22|  6|         1|        46|2013/11/05T04:36:05.651|    0.000|
  [___CTYSEA__]| 61.147.103.134|    0|  778|  1|         1|        68|2013/11/05T04:36:05.653|    0.000|


Site/host counts
All_Sites: 1 (100.00%)
CTYSEA: 1 (100.00%)

rwfind: results saved in rwfind_201311091122_20081.txt rwfind_201311091122_20081_anon.txt
```

Fig. 6.3: Confirmation of network flows related to suspect CIDR block

browser plugin confirms that it does (including showing a record of the search for the suspicious CIDR block from a previous step).

5. The analyst can then search for the same information, this time using a dashboard portal. Using the dashboard user interface, a search is initiated for the IP address found in the network flow report.
6. The search results can be saved to a comma-separated value (CSV) file for further manual processing.

At the end of these steps, the analyst knows more about whether any PRISEM participants had any interaction with these suspect hosts, but these interim results are not integrated into a single report, the contextual knowledge embodied in one part of the system is not carried over into output of another part, and there is no qualification of the events that were identified. Were these scanning attempts blocked (meaning low relevance for response) or were there actual flows that would lead to a conclusion of compromise of any assets (meaning high relevance for response)?

If the steps in the workflow process are too numerous, too manually intensive, and too cumbersome, an analyst is slowed down and rendered less effective or limited in their ability to adequately respond. They may waste time, or may not complete the task, allowing attackers to slip past. If the user must log in to a portal and initiate the process by cutting/pasting individual IP addresses, and pointing/clicking on a *Run!* button, the process will only happen when the human is there to initiate it. Automating these tedious and repetitive tasks, and scheduling some common tasks to be run automatically so the results are waiting to be viewed, frees up the analyst to focus on the hard problems that require human intelligence. This is the only way to increase the velocity of the defender closer to that of the attacker as described by Col. John Boyd in his *OODA Loop* – Observe, Orient, Decide, and Act – construct. *[Boy08]*, *[Ric09]*

6.2 Mission Operations Scenarios

In the following subsections, we will look at some common workflow processes involving IOCs in an operational context. There are three primary use cases of workflows that the DIMS system must serve include the processing of IOCs:

1. Sent into the system in a semi-automated manner;
2. Entered manually in response to external activities (e.g., collaboration in closed, vetted, trust communities, from information passed along from law enforcement, etc.), and;

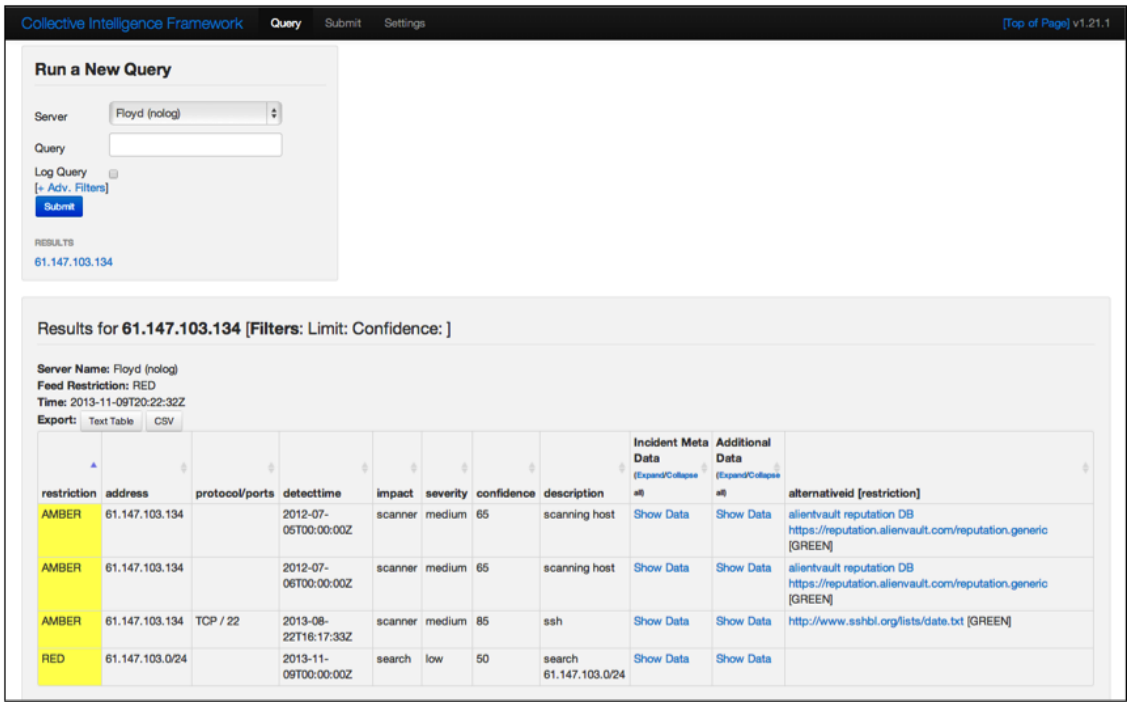

Fig. 6.4: Output of full CIF query via browser plugin

3. As discovered in the iterative and recursive steps taken by an analyst as part of the network forensic process. These use cases parallel the MITRE STIX uses cases UC1, UC3, and UC4 described in Section *Description of needed changes*. Each of these use cases will be described as a separate Mission Operations Scenario.

6.2.1 Automated IOC sharing

Automated sharing of IOCs is not as simple as someone sending an IOC file, which is implicitly acted upon as if it were a request to go search events for some previous period of time and immediately return a report. A human being must validate the results for accuracy and adherence to information sharing policies, approve of the result, and manually release the file to outside parties (possibly after redacting some of the information in the report). This means that even if the first task of performing a historic search is fully automatic, there must be a mechanism for alerting someone that the report is ready for review, multiple automated and asynchronous query results must be queued until they have all been processed, and specific reports must be chosen, analyzed, and released at the appropriate time to the appropriate parties.

There are actually two sub-use cases for automated IOC sharing (one an external-to-internal sharing followed by a reciprocal return internal-to-external sharing, and the other an internal-to-external sharing). Both have privacy sensitivities that require anonymization and controlled release of information.

The first is the situation where US-CERT will be sending de-classified IOCs to the PRISEM system in the form of STIX files, *[The12]* to determine if known malicious activity seen at the federal level is also being seen at the SLTT government level. This is automated input and manual (i.e., vetted and approved) output going back up to the federal level. (Other organizations, such as Microsoft's MAPP program, are similarly being established to share IOCs using STIX, *[Blu13a]*, *[Blu13b]* so STIX packages will become a general input mechanism. An example [abbreviated] STIX file that holds IP addresses and CIDR blocks extracted from a CIF database for use as a *watchlist* is shown in Figure *Example watchlist in STIX format*.)

The second is automated determination of the *sources* of confirmed malicious activity seen at the SLTT level that are collected on a daily basis and prepared for sharing with federal law enforcement and counter-intelligence agents to

Sheets | Charts

	A	B	C	D	E	F
1	COUNT	SOURCE	DESTINATION	DEVICE_VENDC	CUSTOM3	Percent %
2	2	61.147.103.134	12.1	WatchGuard	Firewall	0.03636364
3	3	61.147.103.134	12.1	WatchGuard	Firewall	0.05454545
4	2	61.147.103.134	12.1	WatchGuard	Firewall	0.03636364
5	2	61.147.103.134	12.1	WatchGuard	Firewall	0.03636364
6	1	61.147.103.134	12.1	WatchGuard	Firewall	0.01818182
7	2	61.147.103.134	12.1	WatchGuard	Firewall	0.03636364
8	1	61.147.103.134	12.1	WatchGuard	Firewall	0.01818182
9	1	61.147.103.134	12.1	WatchGuard	Firewall	0.01818182
10	1	61.147.103.134	12.1	WatchGuard	Firewall	0.01818182
11	1	61.147.103.134	12.1	WatchGuard	Firewall	0.01818182
12	1	61.147.103.134	12.1	WatchGuard	Firewall	0.01818182
13	1	61.147.103.134	12.1	WatchGuard	Firewall	0.01818182
14	2	61.147.103.134	12.1	WatchGuard	Firewall	0.03636364
15	3	61.147.103.134	12.1	WatchGuard	Firewall	0.05454545
16	2	61.147.103.134	12.1	WatchGuard	Firewall	0.03636364
17	1	61.147.103.134	12.1	WatchGuard	Firewall	0.01818182
18	1	61.147.103.134	12.1	WatchGuard	Firewall	0.01818182
19	1	61.147.103.134	12.1	WatchGuard	Firewall	0.01818182
20	2	61.147.103.134	12.1	WatchGuard	Firewall	0.03636364
21	1	61.147.103.134	12.1	WatchGuard	Firewall	0.01818182
22	1	61.147.103.134	146.	WatchGuard	Firewall	0.01818182
23	1	61.147.103.134	146.	WatchGuard	Firewall	0.01818182
24	1	61.147.103.134	146.	WatchGuard	Firewall	0.01818182
25	1	61.147.103.134	146.	WatchGuard	Firewall	0.01818182
26	1	61.147.103.134	146.	WatchGuard	Firewall	0.01818182
27	1	61.147.103.134	146.	WatchGuard	Firewall	0.01818182
28	1	61.147.103.134	146.	WatchGuard	Firewall	0.01818182
29	1	61.147.103.134	146.	WatchGuard	Firewall	0.01818182
30	1	61.147.103.134	146.	WatchGuard	Firewall	0.01818182
31	1	61.147.103.134	146.	WatchGuard	Firewall	0.01818182
32	2	61.147.103.134	146.	WatchGuard	Firewall	0.03636364
33	3	61.147.103.134	146.	SecureComputi	Firewall	0.05454545
34	1	61.147.103.134	65.2	CISCO	Firewall	0.01818182
35	1	61.147.103.134	65.2	CISCO	Firewall	0.01818182
36	1	61.147.103.134	OUTSIDE:146.	Cisco	Firewall	0.01818182
37	3	61.147.103.134	outside:206.17	Cisco	Firewall	0.05454545
38	3	61.147.103.134	outside_main	CISCO	Firewall	0.05454545
39						
40						

Fig. 6.5: PRISEM portal CSV output

```
<stix:STIX_Package id="stix:guid-ea022ea8-5a37-11e3-b9db-e4ce8f3f30f8"
version="1.0.1" xmlns:CodeObj="http://cybox.mitre.org/objects#CodeObject-2"
xmlns:ioc-tr="http://schemas.mandiant.com/2010/ioc/TR/"
[...deleted...]
xmlns:xlink="http://www.w3.org/1999/xlink">
    <stix:STIX_Header>
        <stix:Description>PRISEM Watchlist</stix:Description>
    </stix:STIX_Header>
    <stix:Indicators>
        <stix:Indicator id="stix:guid-ea032647-5a37-11e3-9464-e4ce8f3f30f8"
         xmlns:indicator='http://stix.mitre.org/Indicator-2'
         xsi:type='indicator:IndicatorType' negate="false">
            <indicator:Title>Indicator: PRISEM Watchlist</indicator:Title>
            <indicator:Description>An indicator package containing a watchlist</indicator:Description>
            <indicator:Observable id="example:Observable-b6759198-0657-4efc-ab7f-bc0b6c7979da">
                <cybox:Observable_Composition operator="OR">
                    <cybox:Observable id="example:Observable-84c811fa-e494-43fe-9c7c-ccb5421db5f1">
                        <cybox:Object id="example:Address-099ec48a-0664-4348-8bba-248b0b49b1c5">
                            <cybox:Properties xsi:type="AddressObj:AddressObjectType" category="ipv4-addr">
                                <AddressObj:Address_Value>85.233.64.4</AddressObj:Address_Value>
                            </cybox:Properties>
                        </cybox:Object>
                    </cybox:Observable>
                    <cybox:Observable id="example:Observable-c132df60-383c-4a0e-bfa8-68645bcdd4d4">
                        <cybox:Object id="example:Address-fb9a3daa-2271-4541-afec-461a9161cfa6">
                            <cybox:Properties xsi:type="AddressObj:AddressObjectType" category="ipv4-addr">
                                <AddressObj:Address_Value>192.210.200.203</AddressObj:Address_Value>
                            </cybox:Properties>
                        </cybox:Object>
                    </cybox:Observable>
                    <cybox:Observable id="example:Observable-516b4136-9fcd-4c8d-b193-25db9e828137">
                        <cybox:Object id="example:Address-8006e12c-350b-4d3d-98a9-99c0404e3360">
                            <cybox:Properties xsi:type="AddressObj:AddressObjectType" category="ipv4-addr">
                                <AddressObj:Address_Value>219.166.106.109</AddressObj:Address_Value>
                            </cybox:Properties>
                        </cybox:Object>
                    </cybox:Observable>
                    <cybox:Observable id="example:Observable-531595cf-794e-4f3e-87f6-6bcdc5b851d7">
                        <cybox:Object id="example:Address-69aa0e23-00f0-4e9e-ab00-3f72c934dfe8">
                            <cybox:Properties xsi:type="AddressObj:AddressObjectType" category="ipv4-addr">
                                <AddressObj:Address_Value>31.222.200.121</AddressObj:Address_Value>
                            </cybox:Properties>
                        </cybox:Object>
                    </cybox:Observable>
                    <cybox:Observable id="example:Observable-1b25e9c4-9a79-4852-a2f6-11262035e6b4">
                        <cybox:Object id="example:Address-182d6ef4-0ccc-487e-b404-aecd2f0f0485">
```

Fig. 6.6: Example watchlist in STIX format

determine if known cases being investigated by federal agencies involve parties locally. The targets of the attacks (i.e., the sources of the IOCs within the PRISEM participant base) are *not shared*, but only data about the outside *malicious sources*. If federal agents determine that there is a match with an open investigation, they will discretely reach out to a designated contact within the PRISEM system who can assist in reaching out to establish connections with the source (should they chose to make such a connection.) In the future, mechanisms that support *privacy-perserving set intersection* operations based on homomorphic encryption algorithms have proven useful in comparing data sets containing sensitive information without exposing that information to either of the parties involved in the comparison. *[dCT10]* .. _manualentryofiocs:

6.2.2 Manual entry of IOCs

The second case is similar to the external-to-internal sharing use case just described. An analyst or research affiliated with the PRISEM project who may be part of a closed, vetted, trust community, may come to possess information about known or suspected malicious activity derived from investigations performed by another member of said community. That information may be highly sensitive, but also may be highly indicative of targeted activity that has previously escaped the view of the information security vendor and researcher communities, which means it may have bypassed *any and all detective mechanisms* and never triggered an alert within PRISEM's SIEM system (i.e., it is a *false negative*). The analyst would enter data, perhaps in the same way as with the US-CERT IOCs, but processed separately and not queued for potential release to US-CERT. If this check determines there is no evidence of activity within the PRISEM data pool, the analyst is notified. The analyst may optionally chose to enter these indicators into a *watchlist* to alert if/when those indicators are seen in the future (with a note as to why they were put there in the first place, what the suspected activity involved, etc.) This contextual data is best kept in CIF, where it can be correlated with other activity reported by the community in the future. If, on the other hand, there is confirmation that PRISEM participants have been involved in the same activity, the analyst has just performed the first iteration of the next use-case we will consider.

6.2.3 Network Forensic Analysis

The final use case is the most complex, as it involves a series of iterative and recursive queries of available data, going back and forth through time, and extending outward from an initial point to build a network of known hosts involved in various phases (see Figure *Indicator Lifecycle*) of what is known as the *cyber kill chain*. *[HCA11]*

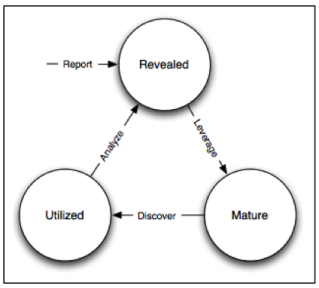

Fig. 6.7: Indicator Lifecycle

The steps described in Section *Generalized Analysis Scenario* and the previous two workflows are repeated, following the process shown in *Indicator Lifecycle*. The discovery and analytic process can refine the understanding of when response actions must be taken, however the deeper an analyst goes using this cycle, the larger the number of discrete files are created in the form of intermediary results and simple output reports. The task of the analyst gets harder and harder to perform as they are buried in related, but unlinked, raw data. This makes it crucial that machine-parseable data be used as both input and output for the steps within each workflow, using a pipeline methodology to take the results of one process and use it in the next step of the process, as well as to attenuate the volume of raw data by applying selective filters to reduce the noise. This is not possible with primitive forms-based browser interfaces that are not designed to maintain and use state (e.g., knowledge gained by the analyst in previous steps) between invocations.

6.3 Mission Support Scenarios

We will now look at some other general Mission Support Scenarios that focus on improving the efficiency of daily communications workflow processes.

6.3.1 Tracking Status of Remediation Efforts

A regular occurrence within the Ops-Trust community is someone reporting a large number of hosts or network autonomous system (AS) numbers that have vulnerable, exploited, or infected computers. The Subject line usually reflects something about the data (e.g., *1.2M NTP amplifiers identified*) Members of the list will read these email messages, extract the list from the body of the message or attached files, process the list (often with a custom script), and do what they can to mitigate the threat within their own network. Some will respond to the email with something like "ACK for AS123, AS456, and AS789".

While these acknowledgement messages are nice, nobody is responsible for tracking them, updating a list with status, etc. It is impossible for one to know, without themselves tracking the entire thread and accumulating the results from all responses, what percentage of the original list of 1.2M items has been mitigated, which ones are left, etc. Such lists are sometimes sent in the body of the message in what is known as a *Cymrufied list* (columns of IP addresses, AS numbers, etc, separated by vertical bar | characters, made popular by Team Cymru. (See Figure *Example "Cymrufied list"*). Sometimes they are Excel spreadsheets attached to the message, or Comma Separated Value (CSV) files. Sometimes people just put a CIDR block in the Subject line of a message. The method is ad-hoc, random, and often requires writing custom scripts to process and extract just the data relevant to one's own network. It is not uncommon to receive a *Cymrufied list* that is placed in a GZIP compressed Unix/Linux `tar` archive file, which is then attached to an email message (necessitating extraction, unpacking the archive, processing the included file with a script, then deleting the `.tar.gz` file, all *manually*.)

The DIMS system will automate this process by supporting the automatic recognition and processing of structured data files either uploaded into the system, attached to email messages, or sent over TAXII or an AMQP message bus. These structured files can then be processed and the context used to track activity (i.e., is this the initial report, an acknowledgement that certain items have been mitigated, etc.) This also allows tracking of the status of mitigation, statistics over time, etc.

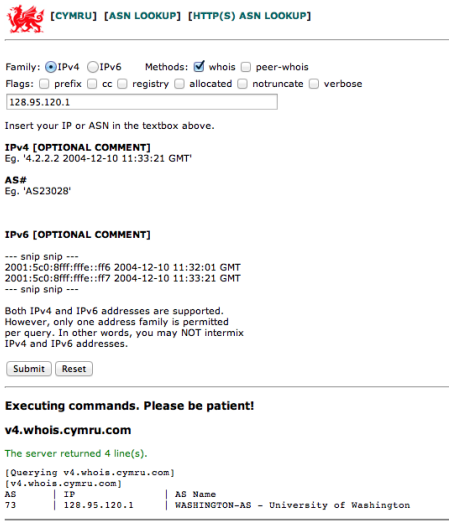

Fig. 6.8: Example "Cymrufied list"

6.3.2 Situational Awareness Through "Identifying Friend or Foe"

When trying to analyze events and alerts in a haystack of data, one method of extracting meaning from the data is to organize it according to facts that are known about the entities that are identified in the haystack of data. A first order of meaning can be derived from taking the end points of connections and categorizing them according to which sets they belong to: known to be a PRISEM participant (a.k.a., *friend*), or known to not be a PRISEM participant.

Table 6.1: Participant identification mapping

CIDR or Domain	Site ID	Participant
156.74.0.0/16	CTYSEA	CTYSEA
.seattle.gov	CTYSEA	CTYSEA
.seattle.wa.gov	CTYSEA	CTYSEA
.seattle.wa.us	CTYSEA	CTYSEA
192.103.189.0/24	PORTTAC	PORTTAC
66.113.101.0/24	PORTTAC	PORTTAC
.portoftacoma.com	PORTTAC	PORTTAC
174.127.160.0/24	COB	BELLWA
12.17.152.0/23	COB	BELLWA
.bellevue.gov	COB	BELLWA
.ci.bellevue.wa.us	COB	BELLWA

Table *Participant identification mapping* illustrates how organizational top-level domains and/or CIDR blocks for a subset of PRISEM participants are mapped to their Site ID strings and chosen anonymization strings (i.e., the label that participant would like to use to mask their internal IP addresses and host names in reports that are shared outside the trust group.) When events are logged, and those logs are ingested into the PRISEM system, they are processed so as to associate them with the site from which they came. Once in the historic log archives, an analyst may search for a specific observable (e.g., *show me all connections to/from a specific suspect IP address.*)

Using this mapping of domains and CIDR blocks to participants, it is possible to identify all records in search results that are associated with any of the PRISEM participants, count how many discrete hosts within each participant site were found, and produce cross-organizational correlation statistics that describe the percentage breakdown of all identified records in the search results. An example of what this process produces can be seen in Figure *Venn diagram of matching/not-matching sets*. In this example, hosts from seven different PRISEM sites were found, with the three most frequent results being in Seattle Childrens Hospital (70.65%), Kitsap County (26.61%), and Port of Olympia (1.38%).

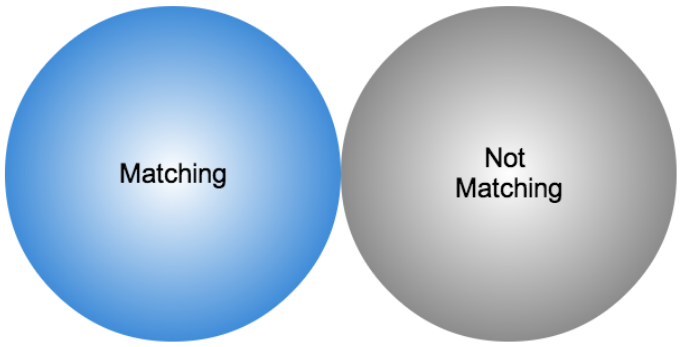

Fig. 6.9: Venn diagram of matching/not-matching sets

Making only one pass over a set of data only allows us to extract IP address and domain names known to be in the map, or not in the map, deriving two non-intersecting sets of entities that are either *matching* and *not matching*. This is depicted graphically with the Venn diagram in Figure *Venn diagram of matching/not-matching sets*.

Without any other information or context about the *not matching* entities that were identified, there is not much that can be deduced about those entities, other than they were involved in connections associated with whatever the analyst was searching for. We can define the results of this pass as identifying *friend* (because we are using a mapping of what constitutes *friend* sites). This is, in fact, how the output of the Cross Correlation service is tagged in Figure

Cross-organizational Correlation of Query Results (Redacted).


```
{"program": "crosscor",
 "date": "Mon Mar 10 18:27:51 PDT 2014",
 "iff": "friend",
 "matching": [
   {"ip4":"[redacted]", "site":"SEA-CHILD"},
   {"ip4":"[redacted]", "site":"SEA-CHILD"},
   {"ip4":"[redacted]", "site":"PORTTAC"},
   {"ip4":"[redacted]", "site":"SEA-CHILD"},
   {"ip4":"[redacted]", "site":"KITSAP"},
   {"ip4":"[redacted]", "site":"KITSAP"},
   {"ip4":"[redacted]", "site":"SEA-CHILD"},
   {"ip4":"[redacted]", "site":"KITSAP"},
   {"ip4":"[redacted]", "site":"SEA-CHILD"},
   {"ip4":"[redacted]", "site":"PORTOLY"},
   {"ip4":"[redacted]", "site":"SEA-CHILD"},
   {"ip4":"[redacted]", "site":"BELLWA"},
   {"ip4":"[redacted]", "site":"SEA-CHILD"},
   {"ip4":"[redacted]", "site":"KITSAP"},
   {"ip4":"[redacted]", "site":"SEA-CHILD"},
   {"ip4":"[redacted]", "site":"CTYSEA"},
   {"ip4":"[redacted]", "site":"CTYSEA"},
   {"ip4":"[redacted]", "site":"BELLWA"},
   {"ip4":"[redacted]", "site":"BELLWA"},
   { ... }
 ],
 "matching_stats": [
   {"site":"PORTOLY","count":"58","percent":"1.38"},
   {"site":"BELLWA","count":"41","percent":"0.98"},
   {"site":"KITSAP","count":"1119","percent":"26.61"},
   {"site":"PORTTAC","count":"7","percent":"0.17"},
   {"site":"KING","count":"5","percent":"0.12"},
   {"site":"SEA-CHILD","count":"2971","percent":"70.65"},
   {"site":"CTYSEA","count":"4","percent":"0.10"},
   {"site":"ALLSITES","count":"4205","percent":"100"}
 ]
}
```


Fig. 6.10: Cross-organizational Correlation of Query Results (Redacted)

Now that we have the list of entities that are not our *friends*, we can make a second pass and add context that will be useful in helping make decisions. Rather than just *known* and *not known,* we can determine, based on information provided by selected authorities to have a certain level of probability of being involved in malicious behavior, that an end point of communication is believed to be hostile (a.k.a., *foe*). The Collective Intelligence Framework accumulates reputation data from sources that the security community deems to be trustworthy in determining which are malicious. If an IP address or domain name occurs in a CIF feed of 65% confidence, then we can assume with 65% confidence that any connections from a PRISEM participant are highly suspicious indicators of malicious activity. If that IP address is not known to any sources that feed CIF, it may or may not be malicious. It could be associated with an *advanced persistent threat* actor who performs targeted attacks and evades the security industry's sandboxes. Or it could be a totally innocent new social network site related to an animal rescue organization. The context and search criteria used by the analyst to get the data being processed holds some clues as to whether the connections are innocent or malicious, and adding context regarding reputation from the security industry and researchers assists even more in making a determination of *innocent* or *malicious* activity.

Figure *Identifying Friend or Foe Based on Reputation Data* illustrates how this second pass works. Starting by identifying those entities that match a mapping of *Friend*, the set of *Not Friend* can then be compared with the set of known malicious entities stored in CIF. Those that are in the intersection of *Not Friend* and *Known to be Bad* by virtue of being found in the CIF database are labeled *Foe*, and the remainder are just *Unknown* at this point. (As an analyst confirms they are actually *Foe*, they should be entered into CIF to allow a positive identification of *Foe* in future queries. This is part of the intelligence gathering process.)

The results of applying the outcome of identifying *Friend* and *Foe* to network flows can be seen in Figure *Graph of all APT1 Related Connections (180 Day Window)* (close-up views of this large graph are found in subsequent figures) These are undirected graphs of connections associated with the set of IOCs released by the FBI in Joint Indicator Bulletin (JIB) #INC260425 in the wake of the release by Mandiant of their *APT1 report*. *[Man13]* Of the 632 IP addresses in the JIB list, it was possible to identify over 7000 flow records associated with 106 hosts on the City of Seattle's network over the previous 180 days. All of those flows were related to just 22 hosts out of the FBI's list

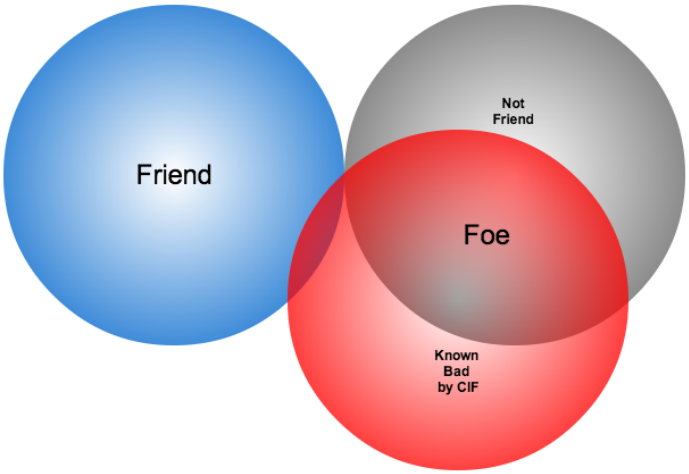

Fig. 6.11: Identifying Friend or Foe Based on Reputation Data

of 632. A search of event logs archived in the PRISEM SIEM identified another three SLTT entities who also had logged events corresponding with indicators on the FBI's list. (In this section, only the City of Seattle network flows are analyzed.)

The cluster in the bottom left of Figure *Patial Graph of APT1 Connection End Points* shows three *Friend* hosts (blue nodes labeled *CTYSEA_nn*) in communication with six JIB-identified (APT1) hosts, only one of which was known by the security industry and made it into the CIF database used by the PRISEM project. Examination of the flows to/from these hosts shows them all to be DNS requests, which is highly indicative of *Fast Flux DNS* for evasion of detection during malware infection. Figure *Connections to a Known Malicious Entity* shows a large number of *Friend* hosts connecting to a known to be malicious APT1 host, while Figure *Connections to an APT1 entity Unknown to CIF* shows an even larger number connecting to an APT1 host that had evaded detection by the security industry and researchers. The context provided by CIF allows rapid triage of the first set, but the lack of known reputation data points to the need to dig deeper and do more thorough analysis of flows and/or perform host-level forensics on the second set of hosts to determine the severity of compromise.

This same process can be applied to textual reports, which could focus on each of the discrete clusters in Figure *Graph of all APT1 Related Connections (180 Day Window)*, including such attributes as: country of origin for non-Friend nodes; AS of origin for non-Friend nodes; Type of activity for *Foe* nodes as known to CIF (including first seen, last seen, etc.); Characterization of identified flows and identified log events (including ports, protocols, start time, duration, etc.).

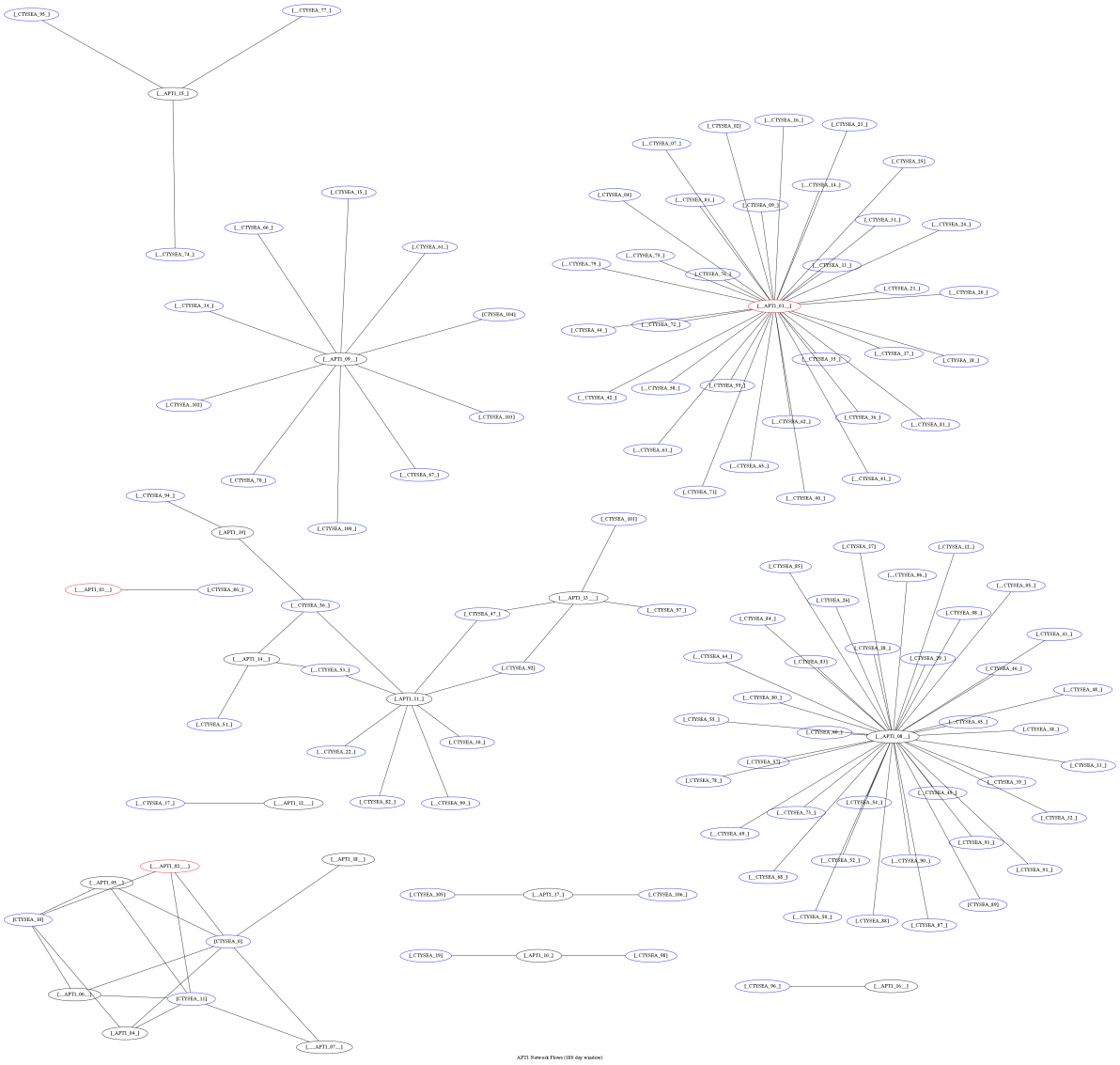

Fig. 6.12: Graph of all APT1 Related Connections (180 Day Window)

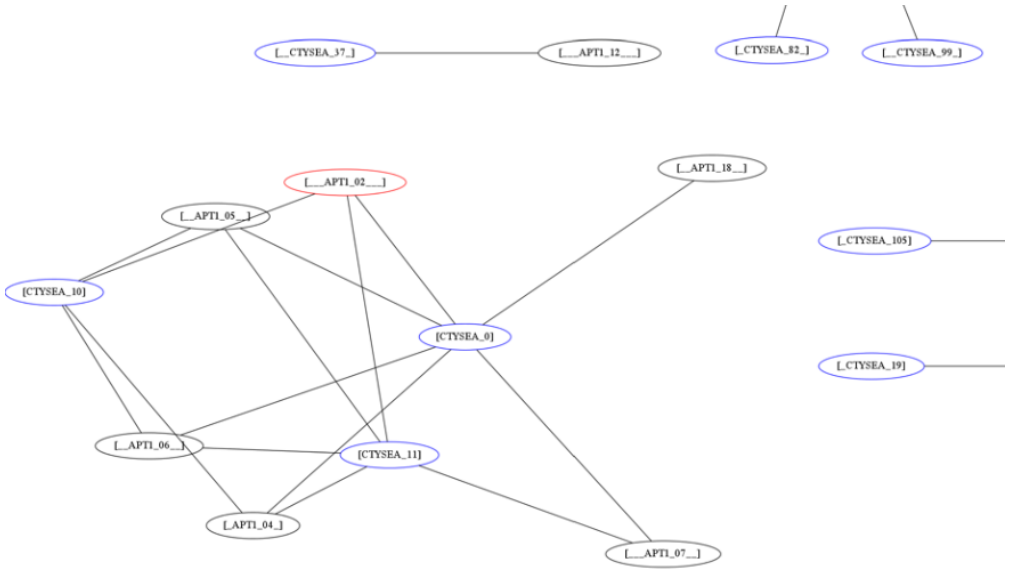

Fig. 6.13: Patial Graph of APT1 Connection End Points

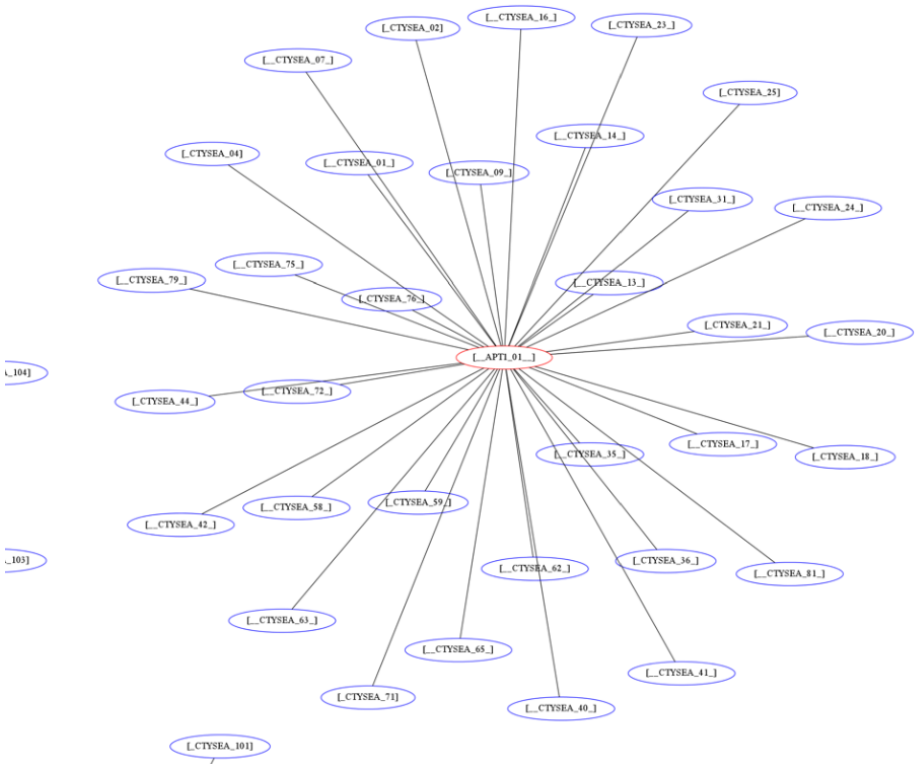

Fig. 6.14: Connections to a Known Malicious Entity

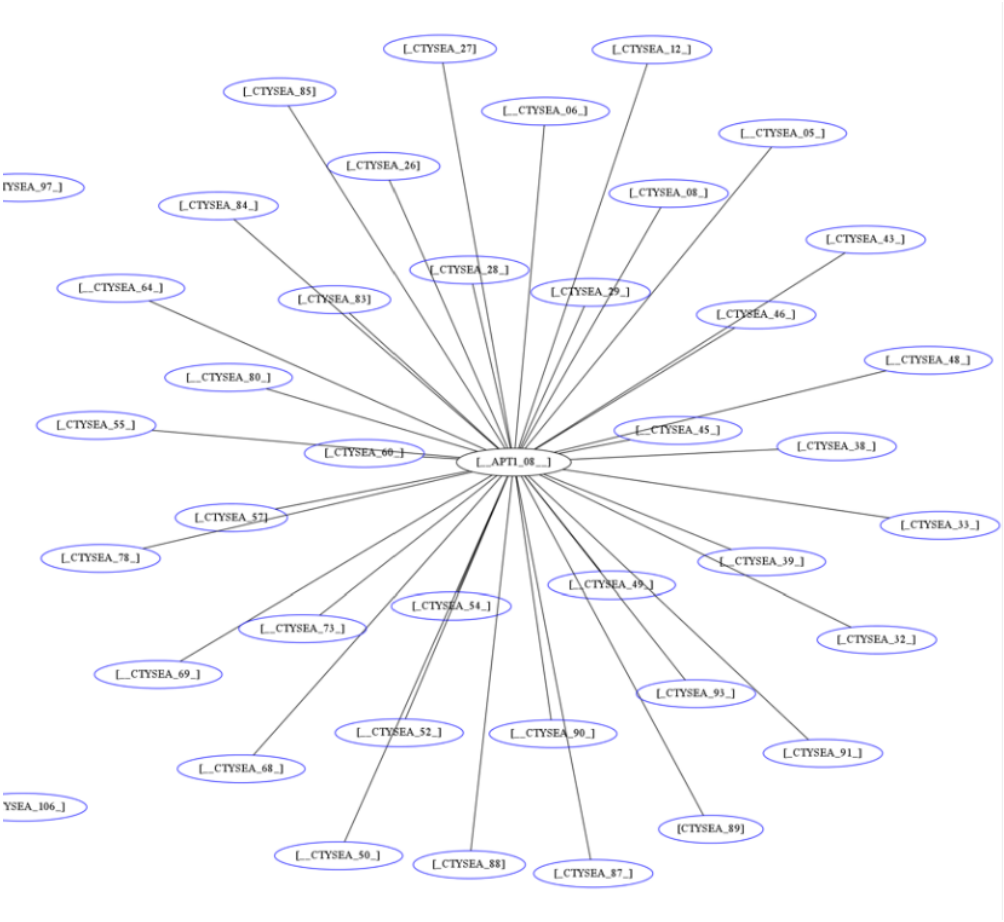

Fig. 6.15: Connections to an APT1 entity Unknown to CIF

CHAPTER 7

Notes

This document is structured on MIL-STD-498, described at A forgotten military standard that saves weeks of work (by providing free project management templates), by Kristof Kovacs. Specifically, this document is modelled on OCD.html.

7.1 Glossary of Terms

Agile A programming methodology based on short cycles of feature-specific changes and rapid delivery, as opposed to the "Waterfall" model of system development with long requirements definition, specification, design, build, test, acceptance, delivery sequences of steps.

Botnets System The name given to the re-implementation of *Einstein 1* technology. See http://web.archive.org/web/20131115180654/http://www.botnets.org/

cron A Unix/Linux service daemon that is responsible for running background tasks on a scheduled basis.

Git A source code version management system in widespread use.

CIFglue "Simple rails app to quickly add indicators to the Collective Intelligence Framework"

Cryptographic Hash

Cryptographic Hashing Algorithm A mathematical method of uniquely representing a stream of bits with a fixed-length numeric value in a numeric space sufficiently large so as to be infeasible to predictably generate the same hash value for two different files. (Used as an integrity checking mechanism). Commonly used algorithms are MD5, SHA1, SHA224, SHA256, RIPEMD-128. (See also http://en.wikipedia.org/wiki/Cryptographic_hash_function).

Einstein 1 A network flow based behavioral and watchlist based detection system developed by University of Michigan and Merit Networks, Inc. for use by US-CERT. The re-implementation is known as the *Botnets System*.

Fusion Center Entities created by DHS to integrate federal law enforcement and intelligence resources with state and local law enforcement for greater collaboration and information sharing across levels of SLTT governments.

GZIP Gnu ZIP (file compression program)

MUTEX Mutual Exclusion (object or lock, used to synchronize execution of independent threads or processes that must share a common resource in an exclusive manner, or to ensure only one copy of a program is running at a time)

NetFlow Record format developed by Cisco for logging and storing Network Flow information (see also SiLKTools).

NoSQL The term for database that does not use the typical table-based relational schema as Relational Database Management Systems (RDBMS)

Ops-Trust (ops-t) Operational Security Trust organization (see http://ops-trust.net/)

Redis A "NoSQL" database system used to store files in a key/value pair model via a RESTful HTTP/HTTPS interface.

SiLKTools A network flow logging and archiving format and tool set developed by Carnegie Mellon's Software Engineering Institute (in support of CERT/CC).

Team Cymru (Pronounced "COME-ree") – "Team Cymru Research NFP is a specialized Internet security research firm and 501(c)3 non-profit dedicated to making the Internet more secure. Team Cymru helps organizations identify and eradicate problems in their networks, providing insight that improves lives."

Tupelo A host-based forensic system (client and server) developed at the University of Washington, based on the Honeynet Project "Manuka" system.

7.2 List of Acronyms

AAA Authentication, Authorization, and Accounting

AMQP Advanced Message Queuing Protocol

AS Autonomous System

ASN Autonomous System Number

CI Critical Infrastructure

CIDR Classless Internet Domain Routing

CIF Collective Intelligence Framework

CIP Critical Infrastructure Protection

CISO Chief Information and Security Officer

COA Course of Action (steps to Respond and Recover)

CONOPS Concept of Operations

CRADA Cooperative Research and Development Agreement

CSIRT Computer Security Incident Response Team

CSV Comma-separated Value (a semi-structured file format)

DIMS Distributed Incident Management System

DNS Domain Name System

DoS Denial of Service

DDoS Distributed Denial of Service

EO Executive Order

HSPD Homeland Security Presidential Directive

ICT Information and Communication Technology

IOC Indicators of Compromise

IP Internet Protocol (TCP and UDP are examples of Internet Protocols)

IRC Internet Relay Chat (an instant messaging system)

JSON JavaScript Object Notation

MAPP Microsoft Active Protections Program

MNS Mission Needs Statement

NCFTA National Cyber-Forensics & Training Alliance

NTP Network Time Protocol (a service exploited to perform reflected/amplified DDoS attacks by spoofing the source address of requests, where the much larger responses flood the victim)

OODA Observe, Orient, Decide, and Act (also known as the "Boyd Cycle")

PISCES Public Infrastructure Security Collaboration and Exchange System

PPD Presidential Policy Directive

PRISEM Public Regional Information Security Event Management (former name, now deprecated - see PISCES)

RBAC Role Based Access Control

RESTful Representational State Transfer web service API

RPC Remote Procedure Call

SCADA Supervisory Control and Data Acquisition

SIEM Security Information Event Management (sometimes referred to as Security Event Information Management, Security Event Monitoring, causing some to pronounce it as "sim-sem".)

SLTT State, Local, Territorial, and Tribal (classification of non-federal government entities)

SOC Security Operations Center

SSH Secure Shell

STIX Structure Threat Information Expression. A standard for information exchange developed by MITRE in support of DHS US-CERT.

TAXII Trusted Automated Exchange of Indicator Information

TCP Transmission Control Protocol (one of the Internet Protocols)

TLP Traffic Light Protocol

TTP Tools, Tactics, and Procedures

UC Use Case

UDP Unreliable Datagram Protocol (one of the Internet Protocols)

WCX Western Cyber Exchange

CHAPTER 8

License

Section author: Dave Dittrich (@davedittrich) <dittrich @ u.washington.edu>

```
Berkeley Three Clause License
=============================

Copyright (c) 2014, 2015 University of Washington. All rights reserved.

Redistribution and use in source and binary forms, with or without
modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this
list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice,
this list of conditions and the following disclaimer in the documentation
and/or other materials provided with the distribution.

3. Neither the name of the copyright holder nor the names of its contributors
may be used to endorse or promote products derived from this software without
specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND
ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED
WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE
DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE
FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL
DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR
SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER
CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY,
OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE
OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
```

Bibliography

[Note1] The PRISEM project is being superceded by a not-for-profit known as the Public Infrastructure Security Collaboration and Exchange System (PISCES). The name *PRISEM* remains in this, and some of the other DIMS documents, but is being replaced as documents are updated.

[Note2] The original portal used by the Ops-Trust community is being re-written and renamed the Trident portal system. It is planned to be released in open source form in Q2-Q4 of 2016.

[Exe03] Executive Office of the President. Homeland Security Presidential Directive 7: Critical Infrastructure Identification, Prioritization, and Protection. http://www.dhs.gov/xabout/laws/gc_1214597989952.shtm, December 2003.

[Fed08] Federal Emergency Management Agency. National Response Framework. http://www.fema.gov/pdf/emergency/nrf/nrf-core.pdf, January 2008.

[Dep13] Department of Homeland Security. Strengthening the Security and Resilience of the Nation's Critical Infrastructure. http://www.dhs.gov/strengthening-security-and-resilience-nation's-critical-infrastructure, August 2013.

[Exe13a] Executive Office of the President. Executive Order No. 13636. http://www.fas.org/irp/offdocs/eo/eo-13636.pdf, February 2013.

[Exe13b] Executive Office of the President. Presidential Policy Directive – Critical Infrastructure Security and Resilience/PPD-21. http://www.whitehouse.gov/the-press-office/2013/02/12/presidential-policy-directive-critical-infrastructure-security-and-resil, February 2013.

[The12] The Mitre Corporation. Standarizing Cyber Threat Intelligence Information with the Structured Threat Information eXpression (STIX). http://makingsecuritymeasurable.mitre.org/docs/STIX-Whitepaper.pdf, 2012.

[Gra12] Will Gragido. Understanding Indicators of Compromise (IOC) Part I. http://blogs.rsa.com/will-gragido/understanding-indicators-of-compromise-ioc-part-i/, October 2012.

[Man11] Mandiant. Using Indicators of Compromise to Find Evil and Fight Crime. http://www.us-cert.gov/GFIRST/presentations/2011/Using_Indicators_of_Compromise.pdf, August 2011.

[Ald12] Jim Aldridge. Targeted Intrusion Remediation: Lessons from the Front Lines. https://www.mandiant.com/blog/black-hat-usa-2012-presentation-targeted-intrusion-remediation-lessons-front-lines/, August 2012. Black Hat USA 2012 Presentation.

[Mic] Microsoft Developer Network. Chapter 3: Workflow and Process. http://msdn.microsoft.com/en-us/library/bb833024.aspx.

[Boy08] John R. Boyd (Col.). Boyd's OODA "Loop" From "The Essence of Winning and Losing", 2008. Available at http://www.d-n-i.net/fcs/ppt/boyds_ooda_loop.ppt.

[Ric09] Chet Richards. Briefings - Colonel John R. Boyd, USAF. http://www.ausairpower.net/APA-Boyd-Papers.html, November 2009.

[Blu13a] Bluehat1. New MAPP Initiatives. http://blogs.technet.com/b/bluehat/archive/2013/07/29/new-mapp-initiatives.aspx, July 2013. BlueHat Blog.

[Blu13b] Bluehat1. MAPP Initiatives Update – Knowledge Exchange Platform. http://blogs.technet.com/b/bluehat/archive/2013/09/16/mapp-initiatives-update-knowledge-exchange-platform.aspx, September 2013. BlueHat Blog.

[dCT10] Emiliano De Cristofaro and Gene Tsudik. Practical private set intersection protocols with linear computational and bandwidth complexity. https://eprint.iacr.org/2009/491.pdf, 2010.

[HCA11] Eric Hutchins, Michael Cloppert, and Rohan Amin. Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains. In 6th Annual International Conference on Information Warfare and Security. Lockheed Martin Corporation, http://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf, December 2011.

[Man13] Mandiant. APT1: Exposing One of China's Cyber Espionage Units. http://intelreport.mandiant.com/Mandiant_APT1_Report.pdf, February 2013.

Index

S

T

U

W