System Requirements and  
Concept of Operations  
for  
  
**FROM LOCAL TO GLOBAL AWARENESS: A DISTRIBUTED INCIDENT MANAGEMENT SYSTEM (DIMS)**

**Table of 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. 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*[[1]](#footnote-1)* 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.

DHS efforts with MITRE to develop information sharing mechanisms based on the Structured Threat Information eXpression (STIX) format, leveraged by a Cooperative Research and Development Agreement between US-CERT and the PRISEM Project, are underway to bring de-classified Indicators of Compromise (IOCs) and Observables from DHS and FBI to regional SLTT government entities for confirmation of involvement of threat actors of national interest. As sharing of these IOCs and 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 take this semi-automated sharing of structured threat information, building on the success of the PRISEM project and leveraging 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 used by 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.

**Revision Summary** (if applicable): **NA**

**SECTION 1: CAPABILITY NEED**

1.1 MNS Required Mission(s) and Need(s)

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 1, 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). (MITRE)

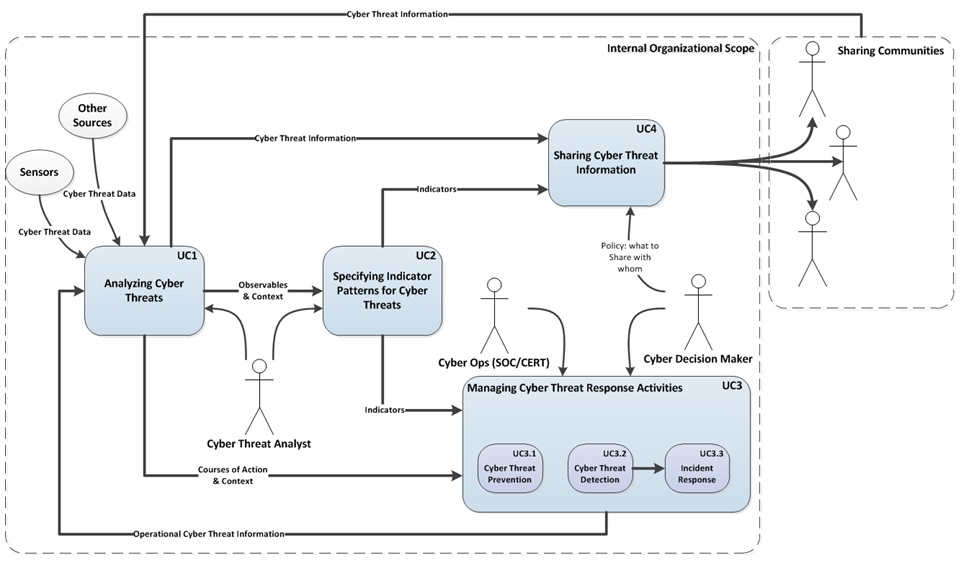


Figure - STIX Use Cases

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.”[[2]](#footnote-4) 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).

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 file hash 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.”[[3]](#footnote-5) 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 awareness reports (SITREPs).

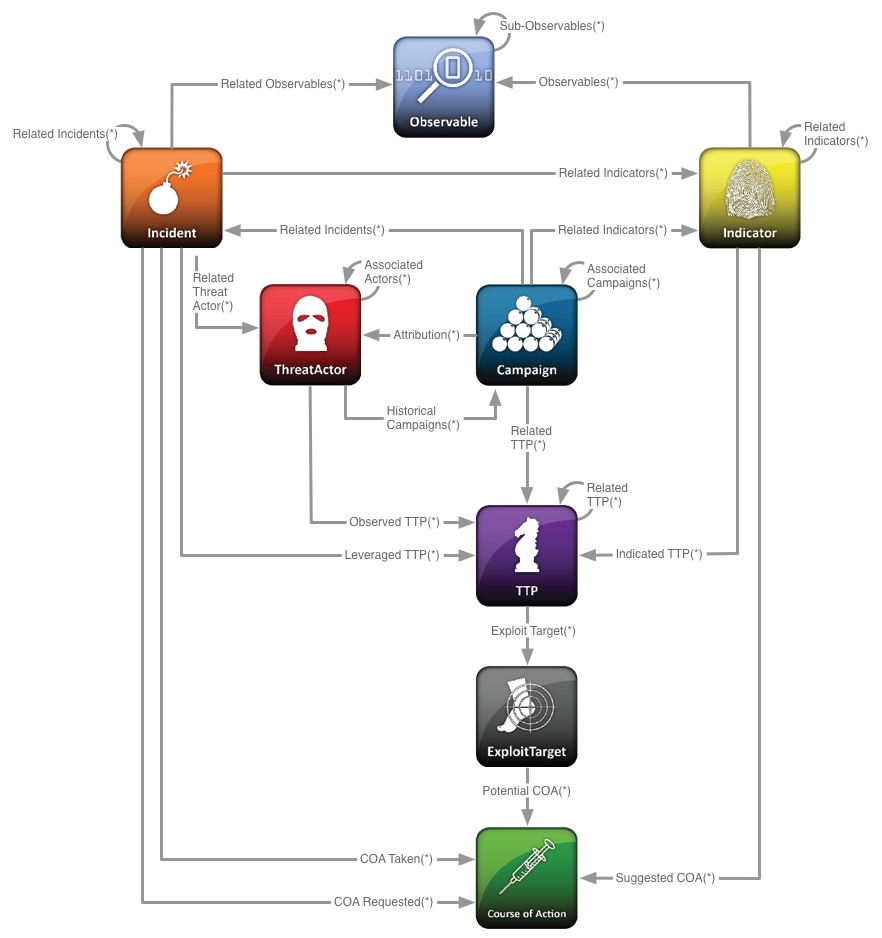


Figure - Relationship of STIX Elements (Source: Bret Jordan, Blue Coat Systems)

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).” (Gragido, 2012) 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 hashes of files], File names, Packer types, Registry keys, Mutexes, DNS strings, and IP Addresses.” (Mandiant, 2011)

IOCs are the lowest-level pieces of evidence used to paint a much larger picture as part of the response and remediation process (Aldridge, 2012). 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.” (Microsoft Developer Network n.d.)

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.

1.2 MNS Capability Gap

The principal high-level gaps in the missions described in the previous section that exist have to do with the availability and affordability of tools that support those missions. Each of them has 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”[[4]](#footnote-6) 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.”[[5]](#footnote-7)

**CURRENT SITUATION**

**SECTION 2: Operations and Support Description**

This section identifies and explains the mission objectives of the DIMS project in terms of users and other stakeholders, and specific operations and support missions.

As the DIMS system relies upon and integrates multiple existing and future open source software components, it will be developed using an Agile programming development methodology (as opposed to the classic “waterfall” development methodology with its sequential processes.) This document, therefore, is a “living document” that will be updated as the project proceeds and as cyclic input/feedback from users and testers is received. Sections to be addressed in future releases of this document are listed as **TBA**.

2.1 Missions (Primary/Secondary)

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 on advancing techniques on the attacker side or the equation. These missions will be described in this section in detail by way of examples of how users and stakeholders will use the DIMS system.

* 1. Users and Other Stakeholders

DIMS is being designed to facilitate trusted information sharing among multiple stakeholder groups, as well as enhancing the ability of a federate group like the PRISEM membership to manage security incidents using a model of event and alert information sharing. A diagram showing a representative subset of these stakeholders and the types of data that will be shared is seen in Figure 3.

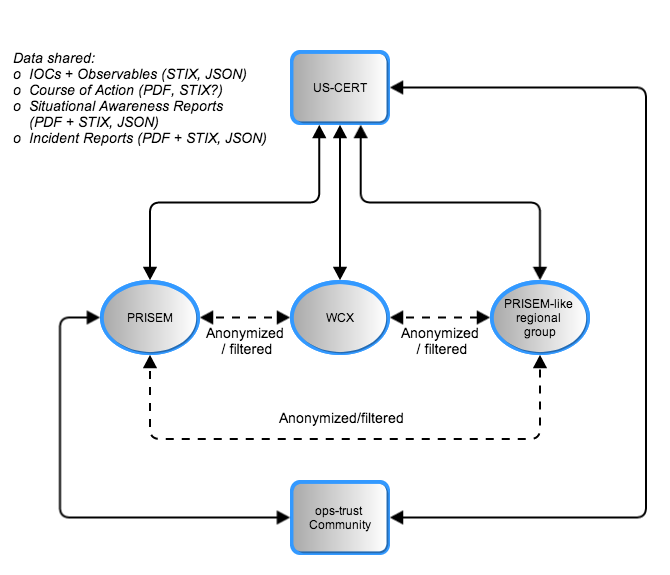
* + 1. 

Figure - Data Flows Between Stakeholders

The full list of stakeholders 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. DHS SOC: Federal government user with some overlapping requirement with PRISEM participants. They operate on a similar model to the threat analyst use case within the PRISEM project, where data across multiple constituents is analyzed for high-level situational awareness as opposed to day-to-day operational needs within individual enterprise units.
9. Policies, Assumptions and Constraints
10. Operational Description

Briefly describe – from a user-oriented perspective – the proposed solution (asset, capability or system), its general employment/operation, and its organizational setting. The operational description includes:

* + 1. Operating Concept (OpCon)

**TBA**

1. Mission Support Description

Mission success depends upon two equally important components: operations and support. While operations is initially described in the MNS (as mission performance), support of the asset or system is first described in the CONOPS. Support is integral to the CONOPS because it is interlaced with operations. Support questions are addressed in a CONOPS. Examples: If a vehicle experiences a significant equipment casualty while underway, it may cease mission execution until the vehicle is repaired. The plan to provide repair support affects the CONOPS; for example the support plan may limit the asset operations to closer-in or more limited operations. If the same personnel performing operational functions on the vehicle also perform repair support functions, as in a minimal support paradigm, they may not have the skills or tools to fix major problems and therefore must avoid hazardous conditions. This may in turn limit their effectiveness in accomplishing the mission.

**TBA**

Since support plays such an important role in this document, the CONOPS working group must include members from the support organizations during the CONOPS draft phase. There are two common models that help describe the support of a system or asset, *The Six Facets of Readiness* or *The Twelve Elements of Logistics* (see chart below). Either may be followed as a guide when writing the mission support description. Briefly describe – from a user-oriented perspective – the concept of mission support for this asset using the *Six Facets of Readiness* or *Twelve Elements of Logistics* framework as a guide. In other words, describe how the lead Component/DHS or asset will support these facets in order to ensure readiness to perform the assigned missions. Topics to discuss include support agency(ies); administrative and medical support; morale, welfare and recreation and work-life considerations; facilities; equipment; configuration management; information technology support; repair/replacement criteria; maintenance levels and cycles; storage, distribution, and supply methods. Identify the different support modes that the asset or system could be in. These support modes later become the titles for the mission support scenarios. For instance, an aircraft might use: Home station, Airborne, Deployed – foreign, Deployed – border patrol facility, Deployed – civilian facility, Depot repair. Information and communications systems might have normal, alerted, high alert, maintenance, etc.

1. Potential Impacts
   * 1. Describe anticipated operational, mission support and other organizational impacts the proposed asset, capability or system will have on the user, acquirer, developer, and support and maintenance organizations. These impacts may include changes in interactions and interfaces with command centers; change in procedures; use of new data sources; changes in quantity, type, and timing of data to be input to the system; changes in data retention requirements; new modes of operation based on peacetime, alert, wartime, or emergency conditions, modification of responsibilities; addition or elimination of responsibilities or positions; need for training or retraining; changes in infrastructure, including facilities and services; and changes in number, skill levels, position identifiers, or location of personnel in various modes of operation. This information allows all affected organizations to prepare for the changes that will be brought about by the new system and to plan for the impacts during development and transition to the new system. The DOTMLPF/R/G/S factor structure from the new Strategic Requirements Planning System should be used to the extent possible to discuss these impacts in a structured manner.

**TBA**

**SECTION 3: SCENARIOS**

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.

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 4 shows a message reporting a suspected network involved in known SSH dictionary scanning, attempting to gain access to insecure accounts.

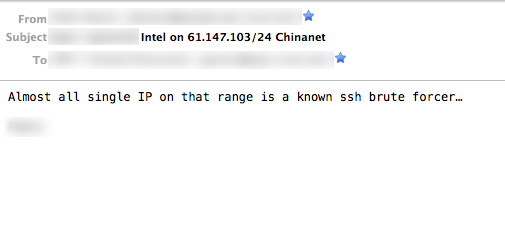


Figure - Email indicator of suspected SSH scanner

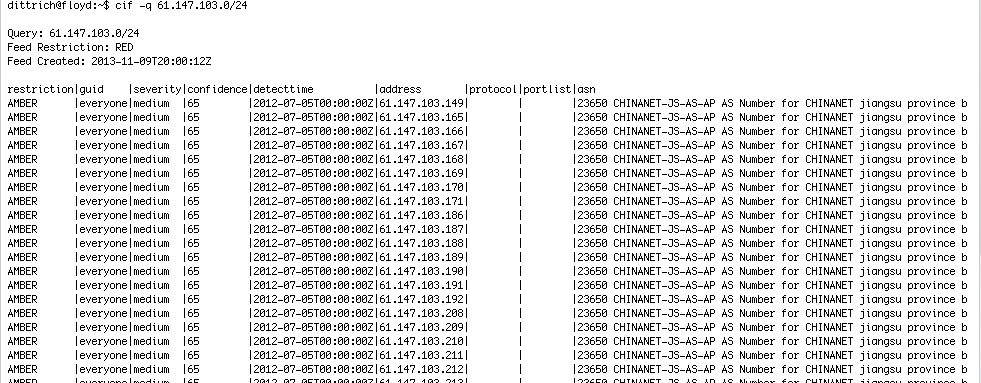
1. The analyst can look in CIF to find what is known about this netblock. From public sources, this network block has been known for a while to be involved in SSH password-guessing attempts. (This screen shot only shows the first few fields from the CIF database.)  
     
   

Figure - Example CIF output for query of suspect CIDR block

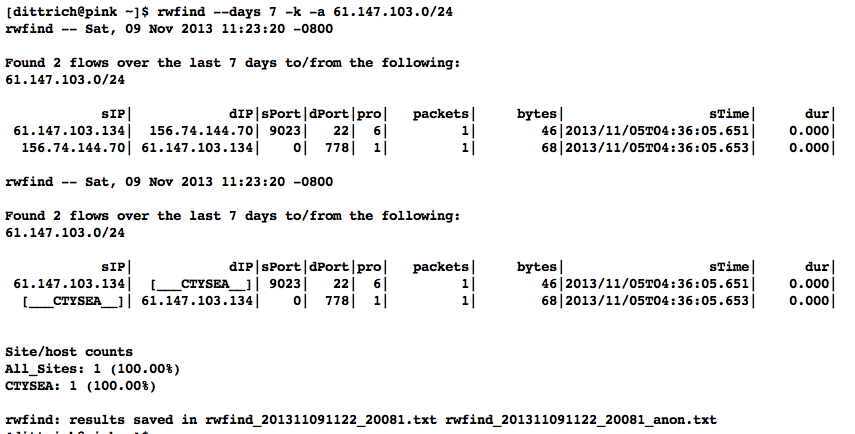
1. The analyst can search historic network flow records to see if there were any recent flows to/from the reported suspect CIDR block. In this case, a seven-day search does turn up some flows. The output is shown here in both raw output form and anonymized using the methods described earlier:  
     
   

Figure - Example output of SiLK flow query for suspect CIDR block

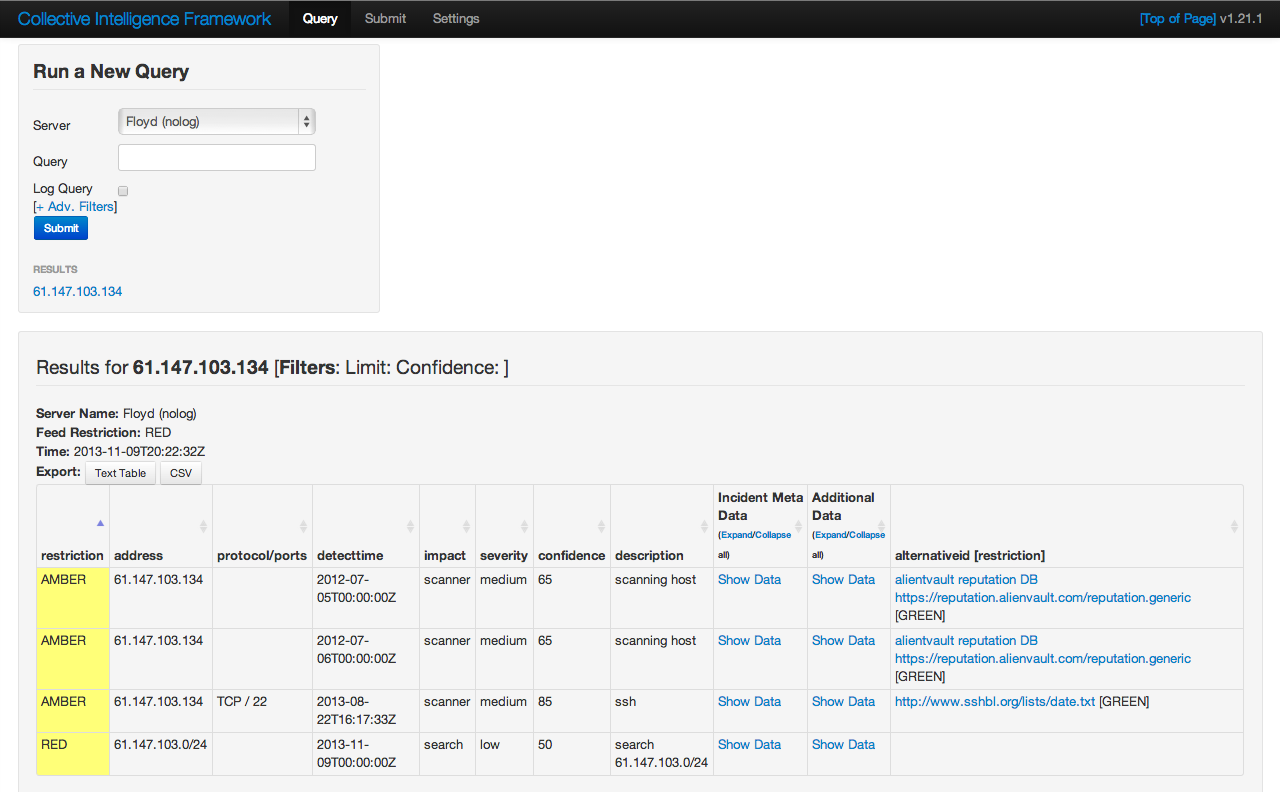
1. The analyst may then query CIF using the web browser interface to see if this specific IP address, seen in the identified flows from the previous step, has any information about it. Figure 7 confirms that it does (including showing a record of the search for the suspicious CIDR block from a previous step).  
     
   

Figure - CIF browser plug-in search results for specific IP address

1. The analyst can then search for the same information, this time using the PRISEM vendor portal. Figure 8 shows the report interface, where a search rule is entered for the IP address found in the network flow report. If the user wanted to search for the entire CIDR block, they would have to enter 256 search terms for each IP address in this interface, since it was not designed to process CIDR blocks (just look for individual IP addresses, or simple substrings thereof).

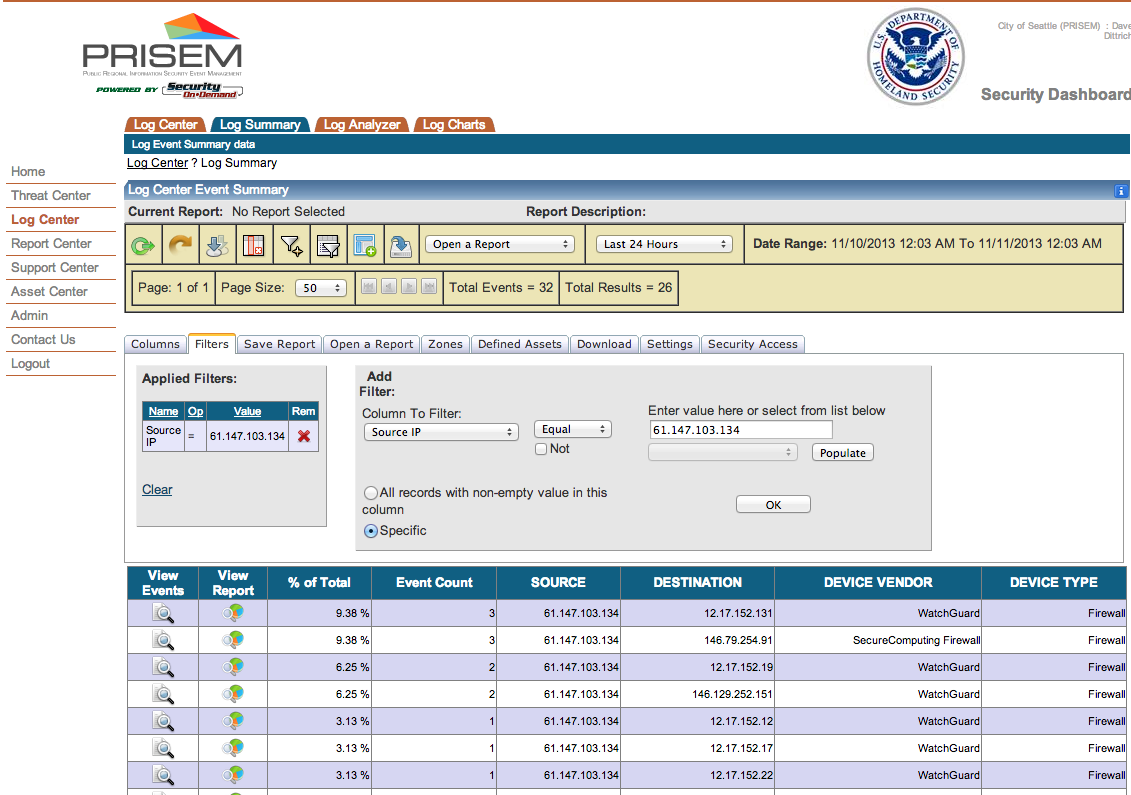


Figure - Vendor Portal (Report Interface)

1. The search results can be saved to a comma-separated value (CSV) file for further manual processing.

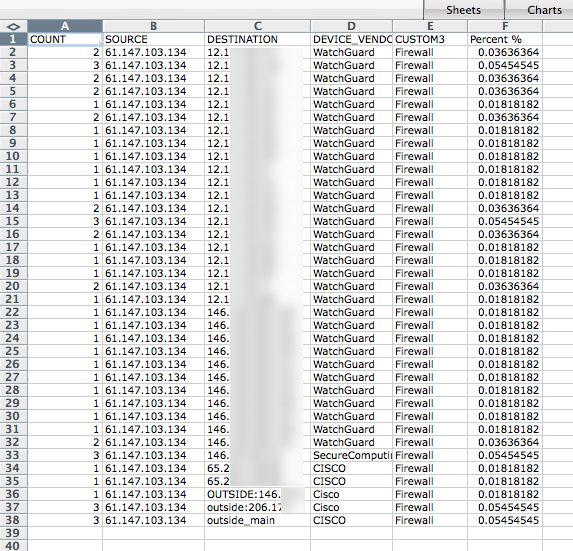


Figure - Saved search results in CSV

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. (Boyd, 2008; Richards, 2009)

1. Mission Operations Scenarios

In this section, we will look at some common workflow processes in an operational context.

There are three primary use cases of workflows that the DIMS system must serve: (1) processing of indicators of compromise sent into the system in a semi-automated manner; (2) processing of indicators of compromise entered manually in response to external activities (e.g., collaboration in closed, vetted, trust communities, from information passed along from law enforcement, etc.), and; (3) as discovered in the iterative and recursive steps taken by an analyst as part of the network forensic process. These use cases parallel those illustrated by MITRE on their STIX web site as UC1, UC3, and UC4. (MITRE 2013) Each of these use cases will be described as a separate Mission Operations Scenario.

* + 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 (Mitre, 2012), 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 (Bluehat1, 2013), 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 10.)



Figure 10 - Example Watchlist in the Form of a STIX Package

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

* + 1. **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.

* + 1. **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 of what is known as the *kill chain* (Hutchins et al, 2011).

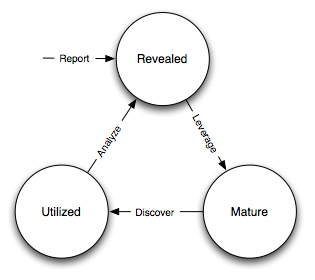


Figure 11 - Indicator Lifecycle

The steps described in Section 3.1 and the previous two workflows are repeated, following the process shown in Figure 11. 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.

This capability is illustrated in Mission Support Scenarios.

* + 1. Functional Capabilities Needed

1. The DIMS system must have the ability to process structured data that is entered into the system in one of several ways: (1) attached to email messages being sent to the Ops-Trust portal (optionally as encrypted attachments); (2) via CIF feed, TAXII, AMQP message bus, or other asynchronous automated mechanism; (3) as uploaded from a user’s workstation via the DIMS dashboard client; (4) via the Tupelo client or other command line mechanism.
2. The DIMS system must have the ability to store additional attributes for each user (such as which CIDR blocks they are responsible for protecting, which top level Domain Name System domains, and/or which high-level activities (e.g., campaigns) they wish to monitor. This capability allows the system to notify the user when there are messages or email threads of interest, and to facilitate providing regular tailored reports or alerts about activity of interest to them. These attributes also support the basis for role-based access controls. This real-time situational awareness capability is one of the most important features that will improve response and reaction time, as it removes the necessity to read and process every single message that flows through the system at a given time, or to manually trigger reports or searches to get situational awareness.
3. The DIMS system must be able to detect when IP addresses or domain names associated with a given set of CIDR blocks or top-level domains are involved, and to trigger one or more workflow processes. This could be to send an alert to a user when some entity they are watching is found in a communication, generate a scheduled report, or trigger some other asynchronous event. It may be to initiate a search of available data so the results can be ready for a user to view when they receive the alert, rather than requiring that they initiate a search at that time and have to wait for the results.
4. The DIMS system must have the ability to process structured data attached to email messages being sent to the Ops-Trust portal (optionally as encrypted attachments), identified by a tag in instant messages or IRC chat (e.g., a URL referencing a data file in a Redis or other “NoSQL” database), as well as detecting when IP addresses or domain names associated with a given set of CIDR blocks or top level domains are found in arbitrary text streams.
5. The DIMS system must be able to keep track of multiple incidents, campaigns, sector-specific threat activity, or other ad-hoc groupings of security information as desired by DIMS users. For example, an analyst may wish to track ZeroAccess trojan activity, CryptoLocker extortion attempts, Zeus or Citadel ACH fraud attempts, etc., possibly over time periods measured in years. Each user may wish to label these associated sets with their own labels, or may want to use a system-wide naming scheme that conforms to an ontology that is more rigorously defined. These sets should be easily shared with other users.
6. The DIMS system should support knowledge acquisition by allowing the user to be told, on login and when they focus on a particular incident or campaign, what new information has been obtained from other users of the system (or the system itself through automated detection and reporting) since the last time the user was reviewing the incident or campaign. Collaboration works best when team members learn from each other, and the asynchronous nature of a multi-user system is such that determining the delta in knowledge since an earlier point in time is difficult to achieve. (This is related to the issue of tracking incoming information in email threads listed earlier.)
7. The DIMS system should summarize any/all aggregate data that any user is presented with sufficient context to quickly understand the data. This includes (but is not limited to): Start and end date and time; Total number of systems within the “friend” population, and how they break down across participants; Total number of systems outside of the “friend” population, and how they break down by country/AS/IP address(es); Total number of systems from the “not-friend” population that are known to be malicious (a.k.a., “foe”), broken down by country/AS/IP address(es). When the number of IP addresses exceeds a certain threshold, they are summarized in aggregate, with a mechanism to dig down if the user so chooses. Similarly, context about what quantity and quality of malicious activity that is known about the “foe” population should also be available for easy access (presented if short, or drill-down provided it too voluminous). This amount and level of detail provides an overall “situational awareness” or scoping of for large volumes of security event data. (The mechanism for such multi-level tabular reports is known as “break”[[6]](#footnote-10) or “step”[[7]](#footnote-11) reports).
8. The DIMS server components must be provisioned, configured, and administered from a single central location and pushed to servers in an automated fashion. Manual configuration and patching of hosts takes too much expert system administration knowledge, incurs too much system administration overhead, and takes too long to recover from outages or system upgrades. The DIMS team will be administering multiple instances of the DIMS system (for development, alpha testing, beta testing, a “production” PRISEM instance for in-field test and evaluation, and potentially 3-5 more instances at other regions (see the Stakeholders section). It will be impossible to manually manage that many deployments with current staffing levels.
9. The systems running DIMS software must support continuous integration of code releases, updating runtime executables, stopping and starting service daemons, etc., in a controlled and repeatable manner. Runtime components must identify the source code release from which they were built in order to track bugs and features across multiple deployments with a regular release cycle measured in weeks (2-4 weeks is anticipated). The system will be built using an Agile coding methodology, responding to user feedback as quickly as possible to ensure maximum usability and scalability.
10. Mission Support Scenarios

The following general Mission Support Scenarios focuses on improving the efficiency of daily communications workflow processes.

* + 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 AS678”. 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.[[8]](#footnote-12) See Figure 12). 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.

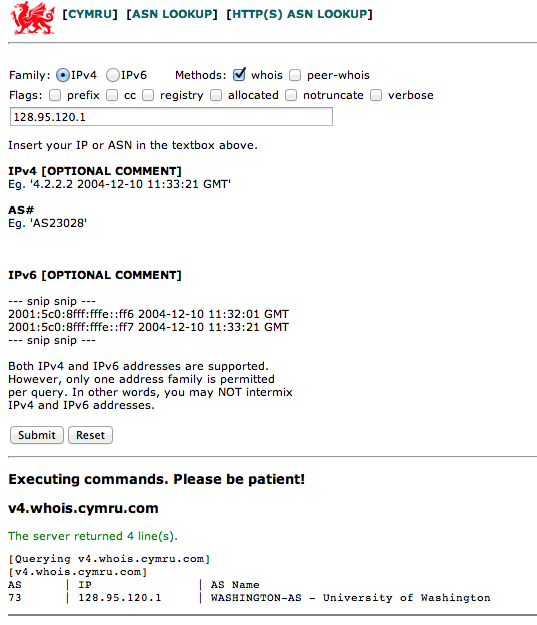


Figure - "Cymrufied list" Example

* + 1. **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.

Figure 13 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.”)

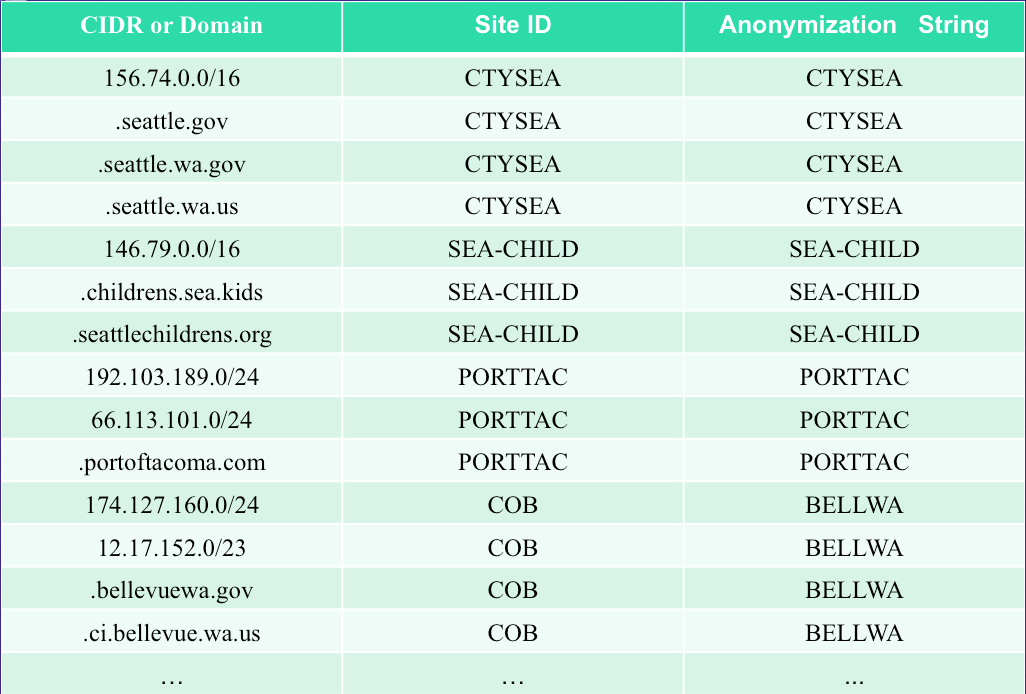


Figure – Partial Participant ID Mapping

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 14. 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%).

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 15. 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 14.

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 - Cross-organizational Correlation of Query Results (Redacted)

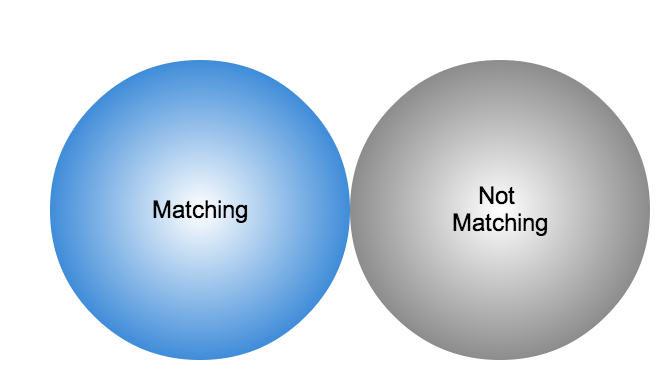


Figure - Simple Identification of End Points of Connections

Figure 16 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.)

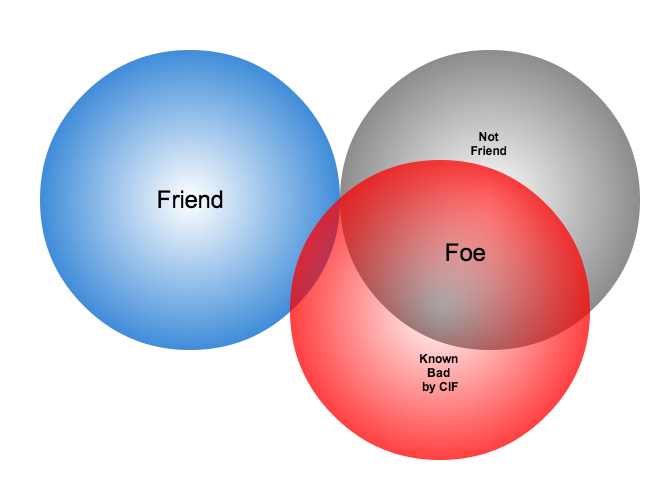


Figure - Identifying Friend or Foe Based on Reputation Data

The results of applying the outcome of identifying “Friend” and “Foe” to network flows can be seen in Figure 17 (close-up views are found in Figure 18 through Figure 20) 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” (Mandiant, 2013). 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 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 18 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”[[9]](#footnote-13) for evasion of detection during malware infection. Figure 19 shows a large number of “Friend” hosts connecting to a known to be malicious APT1 host, while Figure 20 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 17, 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.).

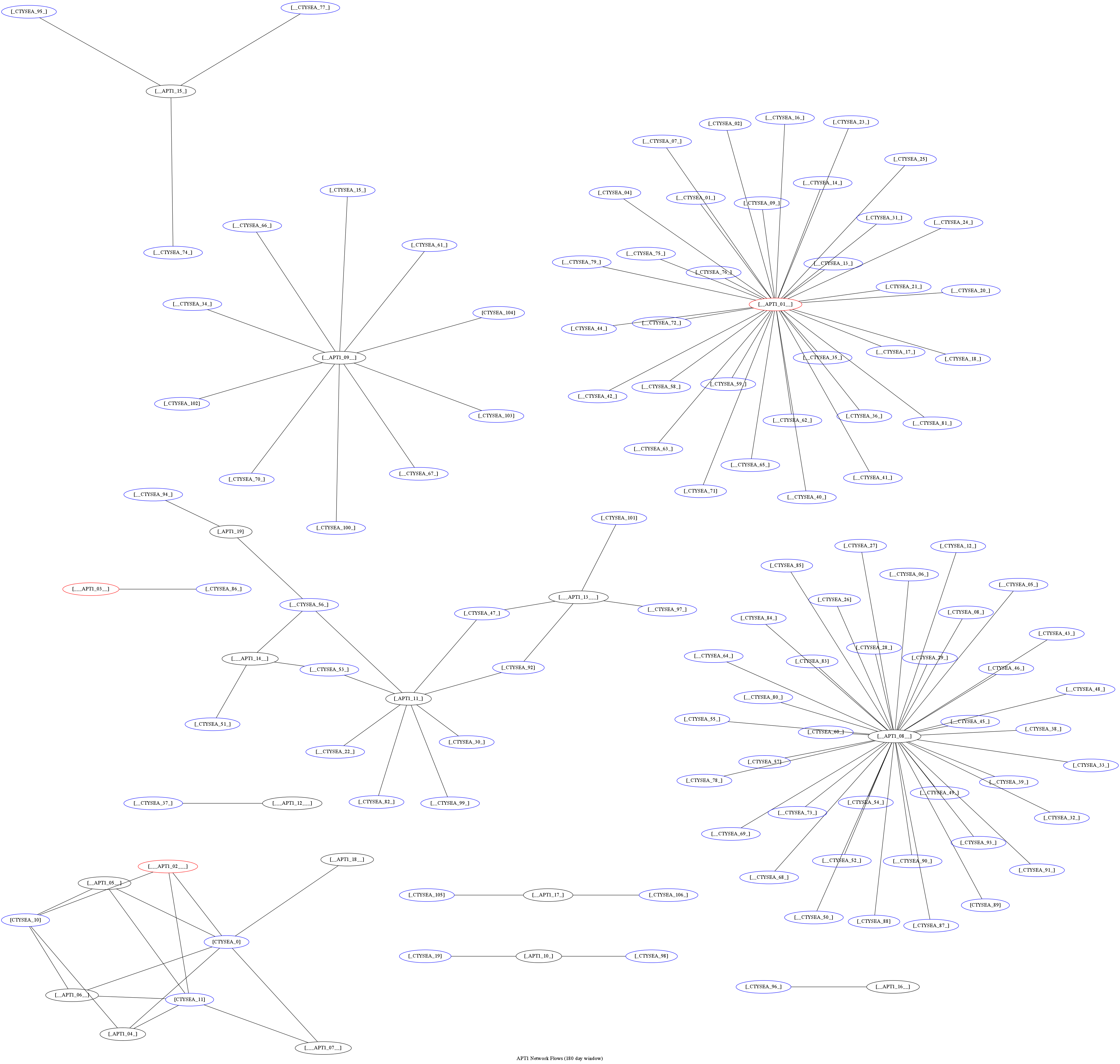


Figure - Graph of all APT1 Related Connections (180 Day Window)

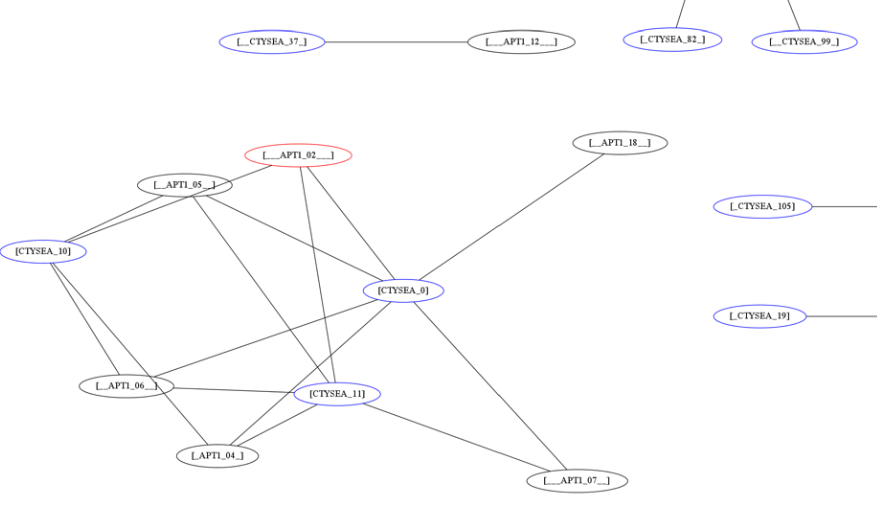


Figure - Partial Graph of APT1 Connection End Points

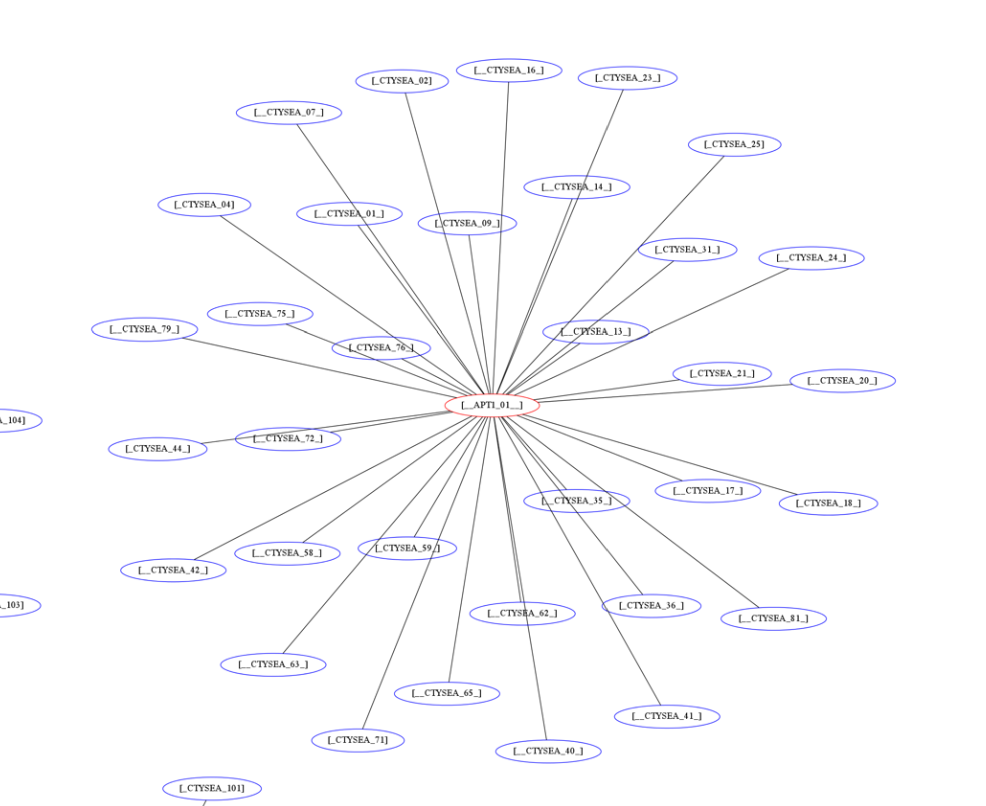


Figure - Connections to a Known Malicious Entity

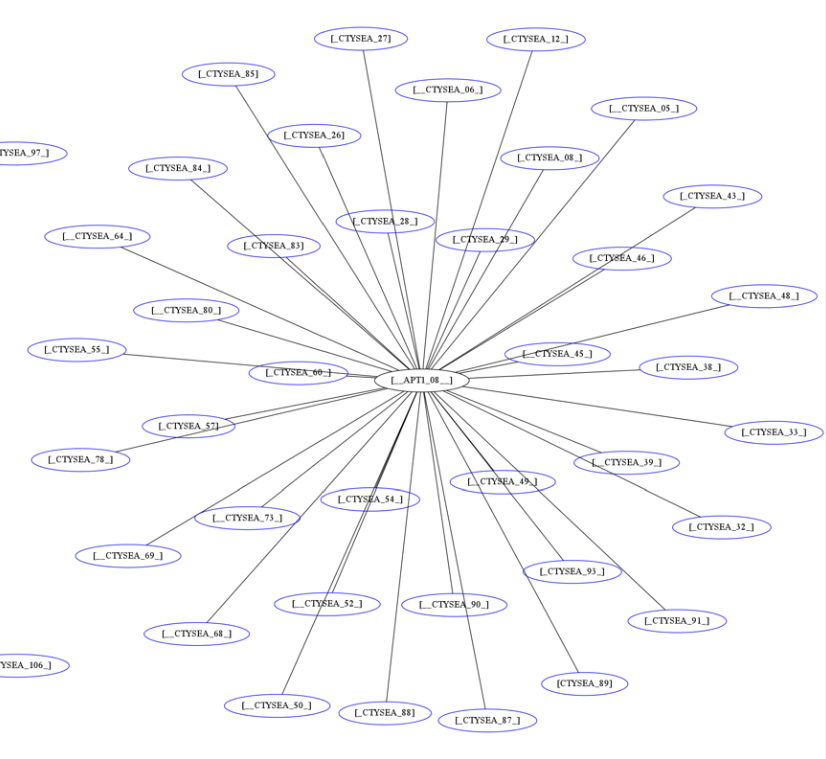


Figure - Connections to an APT1 Entity Unknown to CIF

**4.0 Functional Capabilities**

The functional capabilities of the DIMS system are described in this section as a set of high-level *user stories* as used in Agile programming to provide sufficient detail to developers to allow them to code features that deliver the desired capabilities. These stories will be expanded upon and/or altered based on feedback from stakeholders and prospective users, and will form some of the test and evaluation criteria used to validate the final product.

Users of the system fit into one or more of several categories. They may be analysts, investigators (either private, or law enforcement), system or network administrators, security operations staff, network operations staff, managers, or CISOs. Similar categories are described in research into management of forensic or incident response processes. (Beebe, 2005; Ciardhua ́in, 2004; Hutchins, 2011; Khurana, 2009; Ieong, 2006) In the following user stories, multiple roles may be listed in braces or the general term “user” may be applied. (The stories presented here will need sketched out prototypes of the type of output being described to guide programmers in producing the desired user interface and presentation mechanism.)

1. “As {a security operator, investigator, analyst, CISO} I want to be able to define multiple sets of attributes that the system can then use to inform me about when new data is seen that matches those attributes. Attributes can include anything that might be seen in indicators of compromise, observables, or alerts. (The most basic being IP addresses and/or CIDR blocks, domain names, MD5 or other cryptographic hash values, file names, Registry key settings, etc.)”
2. “As {an investigator, analyst} I want to be able to keep track of cases and campaigns (i.e., groups of related incidents). I want the system to inform me, if I so chose, of any time new data that is determined to be associated with the sets I am tracking comes into the system. For example, if I log in and open a case, I can easily tell which data has been entered into the case since the last time viewed the case. This allows me to stay on top of new evidence or activity that I am investigating.”
3. “As {a security operator, investigator} I want to be told when an email thread or received set of indicators includes systems that I am responsible for securing, ideally pointing out to me those hosts that are involved without requiring that I read the entire thread, extract attachments, write scripts to parse and search data, etc. I want to be given a list of those records that are important, in a format that I can submit directly to query interfaces without having to write scripts to parse and process.”
4. “As {an investigator, analyst} I want to be able to preserve the results of searches, and in some cases the data that was identified while searching, in order to have copies that are subject to expiration and purging from the system. Some investigations may take many months, which could bump up against the data retention period (approximately 12 months, at present).”
5. “As an investigator, I would like to be able to timestamp files I create (i.e., calculate multiple different cryptographic hashes of the contents of files to validate their integrity, associate a timestamp from a trusted time source, then cryptographically sign the result with a private key). This allows validation of the existence of a file at a point in time, who produced the file, and maintenance of a form of “chain of custody” of the contents of the file. To ensure privacy as well as integrity and provenance, the file would first be encrypted (or both cleartext and encrypted files included in the timestamping operation).”
6. “As an {analyst, investigator, security operator}, I would like to be able to get context about “external” hosts that includes what kind of malicious activity has been observed, by whom, starting and ending when, have they been involved in precious incidents I have dealt with, etc. This view could combine a timeline aspect (first seen to last seen time ranges along the X axis), for one or more sources of threat intelligence (discrete items along a non-linear Y axis) with some method of mapping to these external hosts (grouping into AS, etc.). The objective is to quickly associate context about threats within observed flows or logged events.”
7. “As an {analyst, investigator, security operator}, I would like to be able to step through large volumes of output records in a manner that reduces the set of remaining items as quickly as possible. I would like to see related entries visually identified as being part of a common set, and have the ability to select one representative entry, tag it, categorize it as being benign or malicious, then filtering all of the related records out so as to focus on categorizing the remaining records. If the system can remember the tags and automatically apply them when similar records are seen in the future, it will be easier to identify new unknown records that require analytic scrutiny.”
8. “As a system administrator, I would like to have a picture of the operational state of all of the system components that make up DIMS (and related underlying SIEM, etc.) This will allow me to quickly diagnose outages in dependent sub-systems that cause the system as a whole to not function as expected. The less time that it takes me to diagnose the trouble and remediate, the better.”
9. “As a system administrator, I would like to be able to update or reconfigure DIMS subsystem components from a central location (rather than having to log in to each system and copy/edit files by hand). I would like to be assured that those changes are applied uniformly across all subsystem components, and that I have a mechanism to back out to a previous running state if need be to maintain uptime.”
10. “As a {system administrator, security operator}, I would like to know that the DIMS system components are being monitored for attempted access by any of the same malicious actors who are seen to be threatening my constituent users. It is only natural to assume that an attack on any participant site could lead to discovery of the security monitoring system and for that system to be attacked as well, so the system should be monitoring itself using the same cross-organizational correlation features as are used internally.”
11. “As a system administrator, I would like to be able to deal with a breach of the security system in a tactical way. If a user is found to have had a compromise of their account, all access to that user should be disabled uniformly across all system components via the single-signon authentication subsystem. All cryptograph keys should also be revoked. Once the user has been informed and the computer systems they use cleaned, all cryptographic keys, certificates, and password should be updated and re-issued”
12. “As a user of the system, I would like to see the status of any asynchronous queries or report generation requests I have made. It is reasonable for a search through the entire history of billions of events to take some time to complete, but I would like to be able to tell approximately how long I will have to wait. Ideally, the system would keep track of previous requests, the time span and complexity of filtering applied, and to provide a time estimate when a new query is being formulated so as to guide me in deciding what I really need to ask for to get an answer in the time frame I am faced with at the moment.”
13. “As an {analyst, security operator}, I would like to have links to detailed analyses and reports that are available in public sources when a query I have made results in identifying known malware or malicious actors. This way I can more quickly come up to speed on what is (or is not) known about the threat behind the indicators or observables I am dealing with.”
14. “As a {system administrator, security operator, network operator}, I would like to have links to Course of Action steps related to the threats that I identify using the DIMS system. This allows me to not only inform owners or compromised assets that have been identified by the system, but to also give them information about what they need to do, in what order they should take steps, and when/how to preserve evidence in the event that there is criminal investigation ongoing.”
15. “As a user of the DIMS system, I would like the ability to (at any point in time during analysis of an incident or while viewing the situation associated with threats across the user population) produce an anonymized version of the output I am looking at so as to be able to share it with outside entities. The system should anonymize and filter the data according to the policies set by the entities that provided the underlying data, and I should be able to determine the policy for sharing of information (by clearly seeing its tagged TLP sensitivity level). Reports should similarly be tagged appropriately with TLP for the sensitivity level of the aggregate document.”
16. “As a {system administrator, security operator}, I would like to be able to link indicators and observables that come in at the network level (e.g., IP addresses, domain names, URLs) to observables at the host level (e.g., Registry Keys and values, file names, cryptographic hashes of files) and search for those observables to confirm or refute assertions that computers under my authority have been compromised. If I get confirmation, I would then like to preserve evidence and maintain chain of custody for that evidence as easily and quickly as possible.”
17. “As an {analyst, security operator} I would like to be able to start an analysis and annotate data files as I go through the analysis process, trying to derive meaning from what I am seeing in the data, and being able to (at any time seems appropriate) create a reference to the current data set(s) and my view of them so I can pass this reference identifier to another analyst, a CISO, or an investigator, to allow them to take a look at what I am seeing and provide their input. For example, if someone reports a DoS attack directed at SLTT government, and my analysis confirms that such an act can be seen in the PRISEM population, I would like to provide my observations to someone to help investigate targeting, etc., in order to develop a better picture of what is happening. If the result is a determination that a SITREP should be developed and information passed along to federal law enforcement, the updated annotated body of data can then be assembled into a SITREP (using a “break” or “step” reporting format, including both cleartext and anonymized versions for sharing with outside groups) and passed along with little added effort.”

**5.0 Contacts**

David Dittrich (Principal Investigator) [**dittrich@uw.edu**](mailto:dittrich@uw.edu) (206) 685-6473

Mickey Ross (RCES) [mickey.ross@rceservices.net](mailto:mickey.ross@rceservices.net) (425) 971-1615

|  |  |  |
| --- | --- | --- |
| **Version #** | **Date** | **Description** |
| 0.1 | 30 March 2014 | Initial release. |

**6.0 Appendices**

**6.1 Table of Changes**

**6.2 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.

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”[[10]](#footnote-14)

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.

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 name 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.”[[11]](#footnote-15)

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

**6.3 Acronym Listing.**

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”)

PPD – Presidential Policy Directive

PRISEM – Public Regional Information Security Event Management

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

SoD – Security on Demand (PRISEM project support vendor)

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

**6.4 References**

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

N. Beebe and J. G. Clark. A hierarchical, objectives-based framework for the digital investigations process. Digital Investigation, 2(2):147–167, 2005. <http://faculty.business.utsa.edu/nbeebe/pubs/DIP%20Framework%20Journal%20Submission%20v4%20-%20FINAL%20JDI%20author%20copy.pdf>

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

J. 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>

S. O ́. Ciardhua ́in. An Extended Model of Cybercrime Investigations. International Journal of Digital Evidence, 3(1), Summer 2004. <http://www.utica.edu/academic/institutes/ecii/publications/articles/A0B70121-FD6C-3DBA-0EA5C3E93CC575FA.pdf>

D. Dittrich. PRISEM Analyst’s Handbook, December 2013.

D. Dittrich. PRISEM System Administration Handbook, December 2013.

D. Dittrich. Advanced Incident Response Capabilities Supporting Collaborative and Cooperative Responses. Unpublished manuscript, April 2007.

D. Dittrich. On Developing Tomorrow’s “Cyber Warriors”. In Proceedings of the 12th Colloquium for Information Systems Security Education, June 2008. <http://staff.washington.edu/dittrich/misc/cisse2008-dittrich.pdf>

D. Dittrich. On the Development of Computer Network Attack Capabilities. Unpublished manuscript, February 2008. This work was performed for the National Research Council under agreement D-235- DEPS-2007-001.

D. Dittrich. The Conflicts Facing Those Responding to Cyberconflict. In USENIX ;login: vol. 34, no. 6, December 2009. <http://www.usenix.org/publications/login/2009-12/openpdfs/dittrich.pdf>

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.

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

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

M. Hamilton and D. Dittrich. An overview of the Public Regional Information Security Event Management Project, December 2013.

E. Hutchins, M. Cloppert, and R. 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, December 2011. <http://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf>

H. Khurana, J. Basney, M. Bakht, M. Freemon, V. Welch, and R. Butler. Palantir: A Framework for Collaborative Incident Response and Investigation. In IDtrust ’09: Proceedings of the 8th Symposium on Identity and Trust on the Internet, pages 38–51, New York, NY, USA, April 2009. ACM. <http://middleware.internet2.edu/idtrust/2009/papers/05-khurana-palantir.pdf>

R. S. C. Ieong. FORZA - Digital forensics investigation framework that incorporate legal issues. Digital Investigation, 3(Supplement-1):29–36, 2006. <http://www.dfrws.org/2006/proceedings/4-Ieong.pdf>

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

Mandiant. APT1: Exposing One of China’s Cyber Espionage Units, February 2013. <http://intelreport.mandiant.com/Mandiant_APT1_Report.pdf>

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

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

C. Richards. Briefings - Colonel John R. Boyd, USAF, November 2009. <http://www.ausairpower.net/APA-Boyd-Papers.html>

**6.5 List of Figures**

Figure 1 - STIX Use Cases 6

Figure 2 - Relationship of STIX Elements (Source: Bret Jordan, Blue Coat Systems) 7

Figure 3 - Data Flows Between Stakeholders 14

Figure 4 - Email indicator of suspected SSH scanner 17

Figure 5 - Example CIF output for query of suspect CIDR block 17

Figure 6 - Example output of SiLK flow query for suspect CIDR block 18

Figure 7 - CIF browser plug-in search results for specific IP address 18

Figure 8 - Vendor Portal (Report Interface) 19

Figure 9 - Saved search results in CSV 19

Figure 10 - Example Watchlist in the Form of a STIX Package 21

Figure 11 - Indicator Lifecycle 22

Figure 12 - "Cymrufied list" Example 25

Figure 13 – Partial Participant ID Mapping 26

Figure 14 - Cross-organizational Correlation of Query Results (Redacted) 27

Figure 15 - Simple Identification of End Points of Connections 27

Figure 16 - Identifying Friend or Foe Based on Reputation Data 28

Figure 17 - Graph of all APT1 Related Connections (180 Day Window) 29

Figure 18 - Partial Graph of APT1 Connection End Points 30

Figure 19 - Connections to a Known Malicious Entity 30

Figure 20 - Connections to an APT1 Entity Unknown to CIF 31

1. http://web.archive.org/web/20131115180654/http://www.botnets.org/ [↑](#footnote-ref-1)
2. http://stix.mitre.org/about/faqs.html#B1 [↑](#footnote-ref-4)
3. http://stix.mitre.org/about/faqs.html#B2 [↑](#footnote-ref-5)
4. Language used by Wes Young in his presentation on data sharing frameworks at M3AAWG 30th General Meeting in San Francisco, February 20. [↑](#footnote-ref-6)
5. <http://csirtgadgets.org/rfc/getting-started/> [↑](#footnote-ref-7)
6. <http://help.infragistics.com/Help/Doc/Reporting/2012.1/CLR4.0/HTML/Creating%20a%20Multi-Level%20Tabular%20Report.html> [↑](#footnote-ref-10)
7. <http://technet.microsoft.com/en-us/library/cc281376(v=SQL.100).aspx> [↑](#footnote-ref-11)
8. <http://team-cymru.com/> [↑](#footnote-ref-12)
9. See The Honeynet Project. Fast-Flux Service Networks. <http://www.honeynet.org/papers/ff/>, July 2007, and T. Holz, C. Gorecki, and F. Freiling. Detection and mitigation of fast-flux service networks. In Proceedings of NDSS 2008, <http://www.isoc.org/isoc/conferences/ndss/08/papers/16_measuring_and_detecting.pdf> [↑](#footnote-ref-13)
10. <https://github.com/blackfist/CIFGlue> [↑](#footnote-ref-14)
11. <http://www.team-cymru.org/> [↑](#footnote-ref-15)