Federated Threat Data Sharing with the Collective Intelligence Framework (CIF)

Gabriel Iovino (REN-ISAC), Kevin Benton (REN-ISAC), Yoshiaki Kasahara (Kyushu University), Yasuichi Kitamura (APAN)

TIP2013
Honolulu, HI
Funding and Support

Current CIF work is funded by:

  U.S. National Science Foundation OCI SDCI award: “SESv3”,
  Indiana University, and
  REN-ISAC members

Past contributions have also been provided by:

  U.S. Department of Justice, and
  Internet2

In addition, this Inter-Federation project was supported by:

  U.S. National Science Foundation IRNC award: “TransPAC3”
  Asia Pacific Advanced Network (APAN)
What is CIF?

• CIF is a cyber threat intelligence management system. CIF allows you to combine known malicious threat information from many sources and use that information for identification (incident response), detection (IDS) and mitigation (null route).

• The most common types of threat intelligence warehoused in CIF are IP addresses, domains and urls that are observed to be related to malicious activity.

• CIF helps you to aggregate, normalize, store, post process, query, share and produce data sets of threat intelligence.
Aggregation

- CIF can parse data from a wide variety of sources
  - Delimited text, XML, JSON, Zip archives
  - Most other text formats can be parsed as well by providing a regex statement
  - Threading for scalability
  - Simple config files for “new” data sources
Sample Feed Config

severity = medium
confidence = 65
detection = daily
feed = "http://mirror1.malwaredomains.com/files/domains.txt"
impact = 'malicious domain'
source = 'malwaredomains.com'
restriction = need-to-know
alternativeid_restriction = public
guid = everyone

[domains]
values = 'null,null,address,description,alternativeid,null'
delimiter = '\t|\f'
period = daily
Normalization

• Data points are tagged to allow them to be filtered later
• Assessment/Impact
  – What type of threat is this? (botnet infrastructure, phishing domain, spamming host, etc)
• Confidence
  – How confident is the observation? (e.g. is it from an automated process that has false positive?)
• Severity
  – How serious is this threat? Is it part of an active attack? Does seeing this indicator mean you are compromised?
  – Conversely, is this is a known good host that will cause problems if you block it?
Post Processing

• Once data points have been collected by CIF, automated post-processing is performed.
  – Extract domains from URLs
  – Resolve A and NS records
  – Lookup BGP autonomous system information
  – Many more

• Useful for finding related threats
  – e.g. different domains/URLs that point to IP addresses in the same autonomous system
Storage

• v0
  – Data stored as JSON entries in database
  – All server components had to be on same device

• v1
  – Data stored with Google Protocol Buffers
    • ≈70% storage savings
    • 11,491,200 records per day
  – Distributed components provide redundancy and scalability
Producing Data Feeds

• Feeds generated periodically based on criteria
  – e.g. infrastructure related threats with a confidence score of at least 85 from the last 2 days

• Command-line client can output human readable and various machine readable formats
  – Useful for automated generation of firewall rules, URL blacklisting, intrusion detection signatures, etc

• Cached after generation so clients reading the feeds have limited impact on performance
Querying

• Queries can be run on IP addresses, domains, URLs, etc. to see all of the records for a given indicator

• Clients
  – Web Browser Plugin
  – CLI client (Perl)
  – API
  – RequestTracker plugin
# Web Browser Plugin

## Results for 1.1.1.1 [Filters: Severity: medium]

<table>
<thead>
<tr>
<th>restriction</th>
<th>address</th>
<th>protocol/ports</th>
<th>detecttime</th>
<th>impact</th>
<th>severity</th>
<th>confidence</th>
<th>description</th>
<th>Incident Meta Data (Expand/Collapse all)</th>
<th>Additional Data (Expand/Collapse all)</th>
<th>alternativeid [restriction]</th>
</tr>
</thead>
<tbody>
<tr>
<td>need-to-know</td>
<td>safetbrowsing.ru</td>
<td>2012-06-12T00:00:00Z</td>
<td>malicious domain</td>
<td>medium</td>
<td>43.661</td>
<td>rogue</td>
<td>Related Event Show Data</td>
<td>Show Data</td>
<td>malwaresurvival.net/ [public]</td>
<td></td>
</tr>
<tr>
<td>need-to-know</td>
<td>safetbrowsing.ru</td>
<td>2012-06-20T00:00:00Z</td>
<td>malicious domain</td>
<td>medium</td>
<td>43.661</td>
<td>rogue</td>
<td>Related Event Show Data</td>
<td>Show Data</td>
<td>malwaresurvival.net/ [public]</td>
<td></td>
</tr>
<tr>
<td>need-to-know</td>
<td>getpost.in</td>
<td>2012-06-20T00:00:00Z</td>
<td>botnet domain</td>
<td>high</td>
<td>43.661</td>
<td>spyeye</td>
<td>Related Event Show Data</td>
<td>Show Data</td>
<td><a href="https://spyeyetracker.abuse.ch/monitor.php?search=getpost.in">https://spyeyetracker.abuse.ch/monitor.php?search=getpost.in</a> [public]</td>
<td></td>
</tr>
<tr>
<td>need-to-know</td>
<td>vffbgbg.com</td>
<td>2012-06-20T00:00:00Z</td>
<td>botnet domain</td>
<td>high</td>
<td>43.661</td>
<td>spyeye</td>
<td>Related Event Show Data</td>
<td>Show Data</td>
<td><a href="https://spyeyetracker.abuse.ch/monitor.php?search=vffbgbg.com">https://spyeyetracker.abuse.ch/monitor.php?search=vffbgbg.com</a> [public]</td>
<td></td>
</tr>
</tbody>
</table>
Sharing Data

• Groups
  – Each record is associated with a group
  – Each user is issued an API key to use to query the server, which is also associated with a group
  – Queries only return data associated with the same group as the API key
  – Useful for giving other parties access to some data while keeping sensitive data secret
Sharing Data between CIF Installations

• Feeds from a CIF instance can be added as a data source to another CIF instance

• This permits easy data sharing between organizations
  – Only requires one API key for remote organization instead of one for each of its users
  – Organization can remap or override fields based on policies
    • e.g. increase or decrease confidence numbers, change severity, change alternate ID
APAN CIF server installation in APAN Tokyo XP

Yoshiaki Kasahara, Yasuichi Kitamura
kasahara@nc.kyushu-u.ac.jp,
kita@jp.apan.net
Project with REN-ISAC

• Sharing Security Intelligence data between APAN and REN-ISAC
  – They want to have international collaboration (inter-federation information sharing) as their project
  – Currently APAN doesn’t have security response team or something like that...
  – Start from servers around me and Kitamura-san
    • Kyushu Univ. and APAN-JP
What to share?

• Sharing security-related information from intranet is somewhat sensitive...
• There is no such entity collecting security information among APAN community

• It is a test-pilot, so let’s start from easy one

• IP addresses of SSH brute-force attackers
  – Easy to collect from sshd log file
  – Almost no false-positive
  – Sharing is actually beneficial
SSH: Secure Shell

• “SSH (Secure Shell) is intended to replace rlogin and rsh, and provide secure encrypted communications between two untrusted hosts over an insecure network” – ssh(1)

• SSH is secure... only when the user authentication is properly done
  – Encryption doesn’t prevent intruder to use your ID and password (by guessing or other measures)
SSH brute-force attack

• Try to intrude via SSH by using frequently-used ID and password pairs (=dictionary)
  – Fix ID and change passwords
  – Fix password and change IDs
• Mitigation
  – Use a strong password (or one-time password)
  – Use public key authentication
  – Change port of sshd from tcp/22
  – Filter attacking IP addresses
How to filter? (within a host)

• Normally an intruder needs multiple tries with failure
  – The activities are clearly logged like this:

Aug 14 22:39:50 elvenbow sshd[39906]: Invalid user tomcat from 198.101.213.104
Aug 14 22:40:02 elvenbow sshd[39971]: Invalid user tomcat from 198.101.213.104
Aug 14 22:40:47 elvenbow sshd[41218]: Invalid user oracle from 198.101.213.104
Aug 14 22:40:48 elvenbow sshd[41223]: Invalid user test from 198.101.213.104
Aug 14 22:40:56 elvenbow sshd[41235]: Invalid user oracle from 198.101.213.104
Aug 14 22:40:58 elvenbow sshd[41237]: Invalid user test from 198.101.213.104

• Parse these lines and extract attacker’s IP addresses
• Add deny rules to any one of filtering facilities such as iptables, ipfw, hosts.deny, etc
• If the filtering is faster than the correct guess, you win!
Sharing Information

• An attacker cannot attack the whole Internet at once, changing target hosts/networks
  – Other sites can block accesses from the attacker detected at one site, by sharing information

• The more sites, the more protection
Difference?

• There are other tools to protect and share information against SSH brute-force attack

• CIF is a generic tool/framework to handle security intelligence information (not only for SSH attack)
  – Handle various “feeds” (data source)
    • Malware domain blocklist
    • Dragon Research Group (DRG) SSH blacklist
    • bruteforceblocker
    • Spamhaus
    • etc...
Inter-federation information sharing, through algorithmic expressions of policy.
Our Server

• Dell PowerEdge R210 II (1U)
  – Intel Xeon E3-1220 (3.10GHz 4C/4T)
  – 8GB RAM
  – SATA 500GB HDD
  – Ubuntu 11.10
    • Simply we felt it was easier to install and setup than Debian
    • Not sure if it is safe to do-release-upgrade (to 12.04 LTS)
  – Still running CIF 0.01 😞
    • Need to update to CIF 0.05
Our Clients

- The server is pulling SSH attack feeds from...
  - 2 servers in Kyushu University
  - 2 servers in APAN-JP
  - 2 servers in PREGINET (Philippines)
    - He joined during APAN 34th
- Filtering by attack feed is not configured (yet?)
- These feeds are shared with REN-ISAC test server
Topology
Example

- Multiple SSH failures from an IP address detected by a server in Japan
- Results show up on CIF instance in REN-ISAC
Questions?