collective-intelligence-framework

a framework for warehousing intelligence bits

By Russ McRee – ISSA Senior Member, Puget Sound (Seattle), USA Chapter

Prerequisites

Linux for server, stable on Debian Lenny and Squeeze, and Ubuntu v10

Perl for client (stable), Python client currently unstable

As is often the case when plumbing the depths of my feed reader or the Dragon News Bytes mailing list I found *toolsmith* gold. Kyle Maxwell's *Introduction to the Collective Intelligence Framework*[1] (CIF) lit up on my radar screen. CIF parses data from sources such as ZeuS and SpyEye Tracker, Malware Domains, Spamhaus, Shadowserver, Dragon Research Group, and others. The disparate data is then normalized into a repository that allows chronological threat intelligence gathering. Kyle's article is an excellent starting point that you should definitely read, but I wanted to hear more from Wes Young, the CIF developer, who kindly filled me in with some background and a look forward. Wes is a Principal Security Engineer for REN-ISAC whose mission is to aid and promote cyber security operational protection and response within the higher education and research (R&E) communities. As such the tenor of his feedback makes all the more sense.

"The CIF project has been an interesting experiment for us. When we first decided to transition the core components from incubation in a private trust-based community, to a more traditional open-source community model, it was merely to better support our existing community. We figured, if things were open-source, our community would have an easier time replicating our tools and processes to fit their own needs internally. If others outside the educational space benefited from that (private sector, government sector, etc), then that'd be the icing on the cake.

Years later, we discovered that ratio has nearly inverted itself. Now the CIF community has become lopsided, with the majority of users being from the international public and private spaces. Furthermore, the contribution in terms of testing, bug-fixes, documentation contributions and [more importantly] the word-of-mouth endorsements has driven CIF to become its own living organism. The demonstrated value it has created for threat analysts, who have traditionally had to beg-barrow-and-steel their own intelligence, has become immeasurable in relation to the minor investment of adoption.

As this project's momentum has given it a life all its own, future roadmaps[2] *will build off its current success. The ultimate goal of the CIF project is to create a uniform presence of your intelligence, somewhere you control. It'll read your blogs, your sandboxes, and yes, even your email (if you allow it), correlating and digging out threat information that's been traditionally locked in plain, wiki-fied, or semi-formatted text. It has enabled organizations to defend their networks with up-to-the-second intelligence from traditional data sources as well as their peers. While traditional SEMs enable analysts to search their data, CIF enables your data to adapt to your network, seamlessly and on the fly. It's your own personal Skynet. :)"* Readers may enjoy Wes' recent interview on the genesis of CIF, available as a FIRST 2012 podcast.[3]

You may also wish to take a close look at Martin Holste's integration of CIF with his Enterprise Log Search and Archive (ELSA) solution, a centralized syslog framework. Martin has utilized the Sphinx full-text search engine to create accelerated query functionality and a full web front end.[4]

Installing CIF

The documentation found on the CIF wiki[5] should be considered "must read" from top to bottom before proceeding. I won't repeat what's also been said (Kyle's article has some installation pointers too), but I went through the process a couple of times to get it right so I'll share my experience. There are a number of elements to consider if implementing CIF in a production capacity. While I installed a test instance on insignificant hardware running Debian Squeeze, if you have a 64-bit system with 8GB of RAM or more and a minimum of four cores with drive space to grow into, definitely use it for CIF. If you can also install a fresh OS, pay special attention to your disk layout[6] while configuring partition mapping during the Large Volume Manager (LVM) setup. Also follow the postgres database configuration steps[7] closely if working from a fresh install. You'll be changing `ident sameuser to trust in pg_hba.conf` for socket connections. On weak little systems such as my test server, Kyle's suggestion to

1 http://threatthoughts.com/2012/05/07/introduction-to-the-collective-intelligence-framework/.

2 http://www.ren-isac.net/ses/.

3 http://media.first.org/podcasts/FIRST2012-WesYoung.mp3.

4 http://ossectools.blogspot.com/2012/04/accelerating-cif-with-sphinx.html.

5 https://code.google.com/p/collective-intelligence-framework/.

6 https://code.google.com/p/collective-intelligence-framework/wiki/DiskLayout.

7 https://code.google.com/p/collective-intelligence-framework/wiki/PostgresInstall,

Figure 1 – CIF says "here be dragons"

```
high    |65        |2012-06-26T00:00:00Z|193.106.31.68     |       |       |50297 CITONET Centr Informacionnyh Technologii, Ltd.|193.
low     |31.474    |2012-06-26T00:00:00Z|mazilla-update.com|       |       |50297 CITONET Centr Informacionnyh Technologii, Ltd.|193.
low     |17.468    |2012-06-26T00:00:00Z|193.106.31.68     |       |       |50297 CITONET Centr Informacionnyh Technologii, Ltd.|193.
high    |43.661    |2012-06-26T00:00:00Z|mazilla-update.com|       |       |50297 CITONET Centr Informacionnyh Technologii, Ltd.|193.
low     |50        |2012-06-26T00:00:00Z|193.106.31.68     |       |       |50297 CITONET Centr Informacionnyh Technologii, Ltd.|193.
```

Figure 2 – Can I catch a bus out of here?

```
|high    |65        |2012-06-21T00:00:00Z|46.17.96.176      |       |       |49335 NCONNECT-AS Navitel Rusconnect Ltd|46.17.9
|high    |65        |2012-06-21T00:00:00Z|141.105.66.25     |       |       |49335 NCONNECT-AS Navitel Rusconnect Ltd|141.105
|medium  |43.661    |2012-06-21T00:00:00Z|onghdsaleuk.com   |       |       |49335 NCONNECT-AS Navitel Rusconnect Ltd|91.210.
|high    |60.764    |2012-06-21T00:00:00Z|141.105.66.25     |6      |443    |49335 NCONNECT-AS Navitel Rusconnect Ltd|141.105
|high    |43.661    |2012-06-21T00:00:00Z|ns1.microbins.org |       |       |49335 NCONNECT-AS Navitel Rusconnect Ltd|46.17.9
|high    |43.661    |2012-06-21T00:00:00Z|ns2.updatesdns.org|       |       |49335 NCONNECT-AS Navitel Rusconnect Ltd|46.17.9
|medium  |26.532    |2012-06-21T00:00:00Z|91.210.104.97     |       |       |49335 NCONNECT-AS Navitel Rusconnect Ltd|91.210.
|medium  |26.532    |2012-06-21T00:00:00Z|ns2.ipstates.net  |       |       |49335 NCONNECT-AS Navitel Rusconnect Ltd|141.105
|high    |40.157    |2012-06-21T00:00:00Z|ns1.microbins.org |6      | 80    |49335 NCONNECT-AS Navitel Rusconnect Ltd|46.17.9
|high    |40.157    |2012-06-21T00:00:00Z|ns2.updatesdns.org|6      | 80    |49335 NCONNECT-AS Navitel Rusconnect Ltd|46.17.9
|high    |26.532    |2012-06-21T00:00:00Z|46.17.96.246      |       |       |49335 NCONNECT-AS Navitel Rusconnect Ltd|46.17.9
|high    |26.532    |2012-06-21T00:00:00Z|46.17.96.176      |       |       |49335 NCONNECT-AS Navitel Rusconnect Ltd|46.17.9
|high    |26.532    |2012-06-21T00:00:00Z|ns2.ipstates.net  |       |       |49335 NCONNECT-AS Navitel Rusconnect Ltd|141.105
|high    |26.532    |2012-06-21T00:00:00Z|ns2.ipstates.net  |       |       |49335 NCONNECT-AS Navitel Rusconnect Ltd|141.105
|medium  |43.661    |2012-06-21T00:00:00Z|onghdsaleuk.com   |       |       |49335 NCONNECT-AS Navitel Rusconnect Ltd|91.210.
```

update `work_mem` to 512MB and `checkpoint_segments` to 32 in `postgresql.conf` is a good one. The BIND setup[8] is quite straightforward, but again per Kyle's feedback, make sure your forwarder IP addresses in `/etc/resolv.conf` match those you configure in `/etc/bind/named.conf.options`.

From there the install steps on the wiki can be followed verbatim. During the Load Data phase of configuration you may run into an XML parsing issue. After executing `time /opt/cif/bin/cif_crontool -f -d && /opt/cif/bin/cif_crontool -d -p daily && /opt/cif/bin/cif_crontool -d -p hourly` you may receive an error. The `cif_crontool` script is similar to cron, as I hope you've sagely intuited for yourself, where it calls `cif_feedparser` to traverse and load CIF configuration files, then instructs `cif_feedparser` based on the configs. The `error, :170937: parser error : Sequence ']]>' not allowed in content`, crops up when `cif_crontool` attempts to parse the `cleanmx` feed definition in `/opt/cif/etc/misc.cfg`. You can resolve this by simply commenting out that definition. Wes is reaching out to clean-mx.de to get this fixed; right now there are no other options than to comment out the feed.

To install a client you need only follow the Client Setup steps[9], and in your `~/.cif` file apply the `apikey` that you created during the server install as described in CIF Config. Don't forget to configure `.cif` to generate feed as also described in this section.

A final installation note: if you don't feel like spending the time to do your own build, you have the option to utilize a preconfigured Amazon EC2 instance[10] (limited disk space, not production-ready).

Using CIF

You should set the following up, per the Server Install, as a cron job; but for manual reference if you wish to update your data at random intervals, run as `sudo su - cif`:

1. `PATH=/bin:/usr/local/bin:/opt/cif/bin`
2. Pull feed data:
   - `cif_crontool -p daily -T low`
   - `cif_crontool -p hourly -T low`
3. Crunch the data: `cif_analytic -d -t 16 -m 2500` (you can up `–t` and `–m` on beefier systems, but it may grind your system down)
4. Update the feeds: `cif_feeds`

You can run CIF from the command line; `cif –h` will give you all the options, `cif –q <query string>` where query string is an IP, URL, domain, etc., will get you started. Pay special attention to the `–p` parameter as it helps you define output formats such as HTML or Snort.

8 https://code.google.com/p/collective-intelligence-framework/wiki/BindSetup.

9 https://code.google.com/p/collective-intelligence-framework/wiki/ClientSetup.

10 https://code.google.com/p/collective-intelligence-framework/wiki/CIFDemoOnEC2.

```
mazilla-update.com|       |       |              |A   |search       |search mazilla-update.com|                    |
mazilla-update.com|       |       |              |A   |botnet domain|zeus                     |public              |
mazilla-update.com|       |       |193.106.31.68 |A   |search       |search mazilla-update.com|                    |
mazilla-update.com|       |       |ns2.example.com|NS  |search       |search mazilla-update.com|                    |
mazilla-update.com|       |       |ns1.example.com|NS  |search       |search mazilla-update.com|                    |
mazilla-update.com|       |       |193.106.31.68 |A   |botnet domain|zeus                     |public              |
mazilla-update.com|       |       |ns2.example.com|NS  |botnet domain|zeus                     |public              |
```

```
need-to-know|everyone|high    |85        |2012-06-21T00:00:00Z|       |       |756447e177fc3cc39912797b7ecb2f92
need-to-know|everyone|high    |85        |2012-06-26T00:00:00Z|       |       |756447e177fc3cc39912797b7ecb2f92
private     |everyone|low     |50        |2012-06-26T00:00:00Z|       |       |756447e177fc3cc39912797b7ecb2f92
```

```
106.28.0/22|ripencc|UA|botnet infrastructure|zeus                    |
106.28.0/22|ripencc|UA|search               |search mazilla-update.com|
106.28.0/22|ripencc|UA|search               |search mazilla-update.com|
106.28.0/22|ripencc|UA|botnet domain        |zeus                    |
106.28.0/22|ripencc|UA|search               |search 193.106.31.68     |
```

```
6.0/21   |ripencc|RU|botnet infrastructure    |zeus
.64.0/21|ripencc|RU|botnet infrastructure    |zeus
104.0.23|ripencc|RU|malicious domain          |bphoster
.64.0/21|ripencc|RU|botnet url                |zeus binary
6.0/21   |ripencc|RU|botnet domain            |zeus
6.0/21   |ripencc|RU|botnet domain            |zeus
104.0.23|ripencc|RU|malicious infrastructure |bphoster
.64.0/21|ripencc|RU|suspicious nameserver     |stolencardgateway
6.0/21   |ripencc|RU|botnet url               |zeus config
6.0/21   |ripencc|RU|botnet url               |zeus config
6.0/21   |ripencc|RU|botnet infrastructure    |zeus
6.0/21   |ripencc|RU|botnet infrastructure    |zeus
.64.0/21|ripencc|RU|suspicious nameserver     |spyeye
.64.0/21|ripencc|RU|suspicious nameserver     |spyeye
104.0.23|ripencc|RU|suspicious domain         |spammed domain
```

I immediately installed the Firefox CIF toolbar, you'll find details on the wiki under Client | Toolbars | Firefox as it makes queries via the browser, leveraging the API a no-brainer. See WebAPI on the wiki under API. Screen shots included hereafter will be of CIF usage via this interface (easier than manually populating query URLs).

There a number of client examples[11] available on the wiki, but I'm always one to throw real-world scenarios at the tool *du jour*. As ZeuS developers continue to "innovate" and produce modules such as the recently discovered two-factor authentication bypass, ZeuS continues in increased usage by cybercriminals. As may likely be the common scenario, an end user on the network you try desperately to protect has called you to say that they tried to update Firefox via a link "someone sent them" but it "didn't look right" and that they were worried "something was wrong." You run `netstat –ano` on their system and see a suspicious connection, specifically 193.106.31.68. Ruh-roh, Rastro, that IP lives in the Ukraine. Go figure. What does Master Cifu say? Figure 1 fills us in.

I love mazilla-update.com, bad guy squatter genius. You need only web search ASN 49335 to learn that NCONNECT-AS Navitel Rusconnect Ltd is not a good neighborhood for your end user to be playing in. Better yet, `cif –q AS49335` at the command line or drop AS49335 in the Firefox search box.

Figure 2 is a case in point, Navitel Rusconnect Ltd is definitely the wrong side of the tracks.

11 https://code.google.com/p/collective-intelligence-framework/wiki/ClientExamples.

Figure 3 – Mazilla <> Mozilla

```
https://zeustracker.abuse.ch/monitor.php?search=mazilla-update.com
```

```
https://zeustracker.abuse.ch/monitor.php?search=mazilla-update.com
https://zeustracker.abuse.ch/monitor.php?search=mazilla-update.com
```

Figure 4 – CIF hash search

```
|    |malware|zeus binary
|    |malware|zeus binary
|    |search |search 756447e177fc3cc39912797b7ecb2f92
```

ZeuS configs and binaries, SpyEye, stolen credit card gateway, oh my.

This is a good time for a quick overview of taxonomy. Per the wiki, *severity* equates to seriousness, *confidence* denotes faith in the observation, and *impact* is a profile for badness (ZeuS, botnet, etc.).

Our above-mentioned user does show mazilla-update.com in their browser history; let's query it via CIF.

Figure 3 further validates suspicions.

You quickly discern that your end user downloaded `bt.exe` from mazilla-update.com. You take a quick md5sum of the binary and drop the hash in the CIF search box. 756447e177fc3cc39912797b7ecb2f92 bears instant fruit as seen in Figure 4.

Yep, looks like your end user might have gotten himself some ZeuS action.

With a resource such as CIF at your fingertips you should be able to quickly envision value added when using a DNS sinkhole (hello 127.0.0.1) or DNS-BH from malwaredomains.com, where you serve up fake replies to any request for the likes of mazilla-update.com. Bonus! Beefy server for CIF: $2499. CIF licensing: $0. Bad guy fail? Priceless.

In conclusion

Check out the Idea List in the CIF Projects Lab; there is some excellent work to be done including a VMWare appliance, further Snort integration, a Virus Total analytic, and others. This project, like so many others we've discussed in toolsmith, grows and prospers with your feedback and contributions. Please consider participating by joining the CIF Google Group[12] and jumping in. You'll also want to check out the *DFIR Journal*'s CIF discussions,[13] including integration with ArcSight, as well as EyeIS's CIF incorporation with Splunk.[14] These are the same folks who have brought us Security Onion 1.0 for Splunk, so I'm imaging all the possibilities for integration. Get busy with CIF, folks. It's a work in progress but an excellent one at that.

Cheers…until next month. Ping me via email if you have questions (russ at holisticinfosec dot org).

Acknowledgements

—Wes Young, CIF developer, Principal Security Engineer, REN-ISAC

About the Author

Russ McRee leads the incident management and penetration testing functions for Microsoft's Online Services Security team. He advocates a holistic approach to information security via holisticinfosec.org and volunteers as a handler for the SANS Internet Storm Center. Reach him at russ at holisticinfosec dot org or @holisticinfosec.

12 https://groups.google.com/forum/?fromgroups#!forum/ci-framework

13 http://dfirjournal.wordpress.com/2012/04/28/cif-integration-with-arcsight.

14 http://eyeis.net/2012/04/querying-cif-data-from-splunk/