PROCEEDINGS OF SCIENCE

WLCG Security Operations Centres Working Group

David Crooks[*]
University of Glasgow
E-mail: david.crooks@glasgow.ac.uk

Liviu Vâlsan
CERN
E-mail: liviu.valsan@cern.ch

Security monitoring is an area of considerable interest for sites in the Worldwide LHC Computing Grid (WLCG), particularly as we move as a community towards the use of a growing range of computing models and facilities. There is an increasingly large set of tools available for these purposes, many of which work in concert and use concepts drawn from the use of analytics for Big Data. The integration of these tools into what is commonly called a Security Operations Centre (SOC), however, can be a complex task - the open source project Apache Metron (which at the time of writing is in incubator stage and is an evolution of the earlier OpenSOC project) is a popular example of one such integration. At the same time, the necessary scope and rollout of such tools can vary widely for sites of different sizes and topologies. Nevertheless, the use of such platforms could be critical for security in modern Grid and Cloud sites across all scientific disciplines.

In parallel, the use and need for threat intelligence sharing is at a key stage and is an important component of a SOC. Grid and Cloud security is a global endeavour - modern threats can affect the entire community, and trust between sites is of utmost importance. Threat intelligence sharing platforms are a vital component to building this trust as well as propagating useful threat data. The MISP software (Malware Information Sharing Platform) is a very popular and flexible tool for this purpose, in use at a wide range of organizations in different domains across the world.

In this context we present the work of the WLCG Security Operations Centres Working Group, which was created to coordinate activities in these areas across the WLCG. The mandate of this group includes the development of a scalable SOC reference design applicable for a range of sites by examining current and prospective SOC projects & tools. In particular we report on the first work on the deployment of MISP and the Bro Intrusion Detection System at a number of WLCG sites as SOC components, including areas of integration between these tools. We also report on our future roadmap and framework, which includes the Apache Metron project.

International Symposium on Grids and Clouds 2017 -ISGC 2017-
5-10 March 2017
Academia Sinica, Taipei, Taiwan

[*] Speaker.

1. Introduction

The growth of the use of clouds and other virtualised enviroments throughout the Grid and elsewhere is leading to a more challenging task of making sure that contributing Cloud and Grid sites are secure. A key component of this work is in the area of security monitoring. In a similar timeframe, analytics have become a larger part of Grid monitoring at a system and experiment level. We can apply similar concepts to security, which gives rise to the idea of a Security Operations Center (SOC). The purpose of a SOC is to gather relevant security monitoring data from different sources and aggregate that data for use in the detection of security events and acting on them. It is important to have these tools well understood and implemented in a way to best serve the sites and community as a whole. The mandate of the SOC working group is to investigate different models for SOCs and advise the WLCG on best practice from the experience of the group.

2. Framework

Below you can see the architecture diagram for a SOC project called Metron [1], which has grown from a previous project called OpenSOC [2][1]. Metron is the project being used as a reference framework for the working group as a project in active development that captures the long term goals of the working group. As can be seen from this architecture, a SOC consists of components including data ingest, storage, transport, enrichment and alerting. The goal of the WLCG working group is to look at the different components of a SOC and build on these in a staged manner. In this way it is hoped that the experience of contributing sites may be best reflected alongside the needs of the WLCG community as a whole. Indeed, the work of the group has focussed around the development of a minimum viable product - a small set of tools which, when combined, could provide a basic level of protection which was nevertheless highly useful.

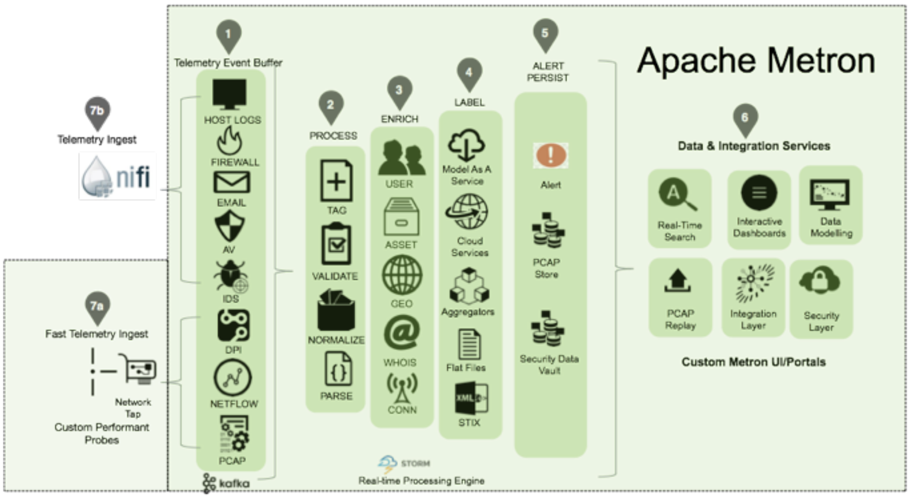

Figure 1: The Metron architecture.

[1] OpenSOC began as the SOC project under CISCO which was released under an open-source licence and subsequently moved to be an incubator project under Apache.

It is important, however, that before discussing this work in more detail, we first discuss the work of the CERN Security Team on the CERN SOC.

3. CERN SOC

In this section we discuss the current status of the CERN Security Operations Center, particularly in the context of this working group.

The architecture diagram for the complete CERN SOC is shown below.

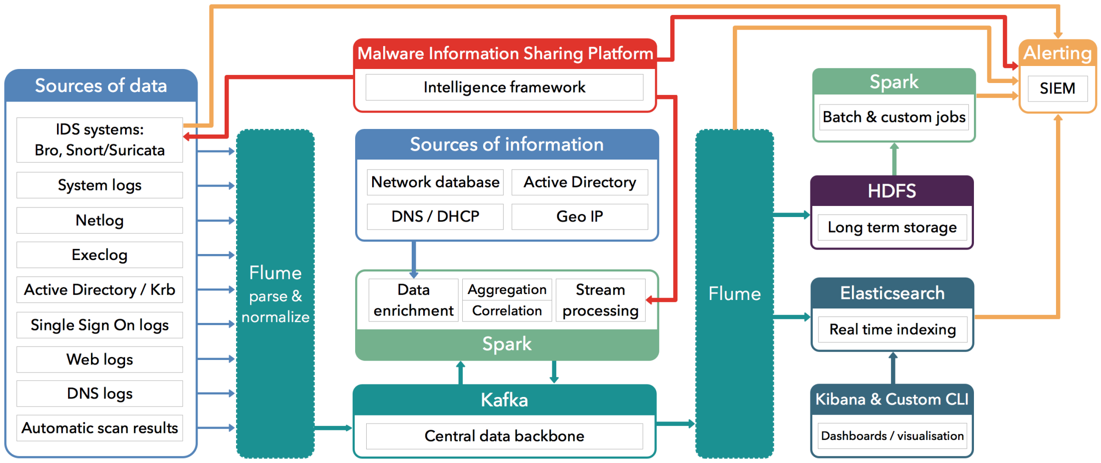

Figure 2: The architecture of the CERN SOC.

A multitude of input sources are used:

- Intrusion Detection Systems:
  - Bro IDS, the primary Intrusion Detection System used at CERN
  - Snort, with a possible future migration to Suricata in the future
- System Logs:
  - General system logs, priority being given to the security sensitive logs (e.g. SSH, PAM authentication, audit logs, etc)
  - Logs are being collected both from Linux and Microsoft Windows systems
- Netlog [3]:
  - Linux kernel module that logs TCP/UDP high lever activity via the following syscalls:
    - ∗ TCP connect: inet_stream_connect
    - ∗ UDP 'connect': inet_dgram_connect
    - ∗ TCP accept: sys_accept(4)
    - ∗ UDP/TCP close: sys_close

        - UDP bind: sys_bind

- Execlog [3]:
    - Linux kernel module that logs all calls to the ’execve’ syscall, effectively tracking all users executions:
- Active Directory and Kerberos logs
- Single Sign On logs
- Web logs
- DNS logs
- Logs of automatic scans (inventory of assets on the network, vulnerability scans, etc.)

3.1 Log transport

Apache Kafka is used as a central data transport layer to which all data is written to and read from. It also acts as a data buffer with the last 24 hours of log data available in the Apache Kafka cluster. Apache Flume is being used both for ingesting data into Kafka from the various sources mentioned above and at the same time to write from Kafka to the various data storage systems, listed below.

3.2 Data analysis

Apache Spark is being used both for streaming and batch data analysis, making use of the data available in Kafka.

3.3 Data storage

Data is being stored in two different storage systems:

- HDFS: For long term storage (one year)
- ElasticSearch: For real time indexing, access and visualisation (data stored in ElasticSearch for three months)

We now look at areas of particular relevance to the working group.

3.4 Intrusion Detection System: Bro

Bro is the primary Intrusion Detection System used at CERN. While Bro is a single process application it comes with support for cluster operations. That mode of operation is achieved by having multiple Bro processes acting as individual workers that are being synchronised through the use of a manager process and of a proxy process. At CERN a Brocade MLXe network router is used as a traffic aggregator and splitter, using symmetrical hashing of the network traffic.

Below you can find the networking diagram showing the way network traffic at CERN is being aggregated and split over to the individual nodes in the Bro cluster.

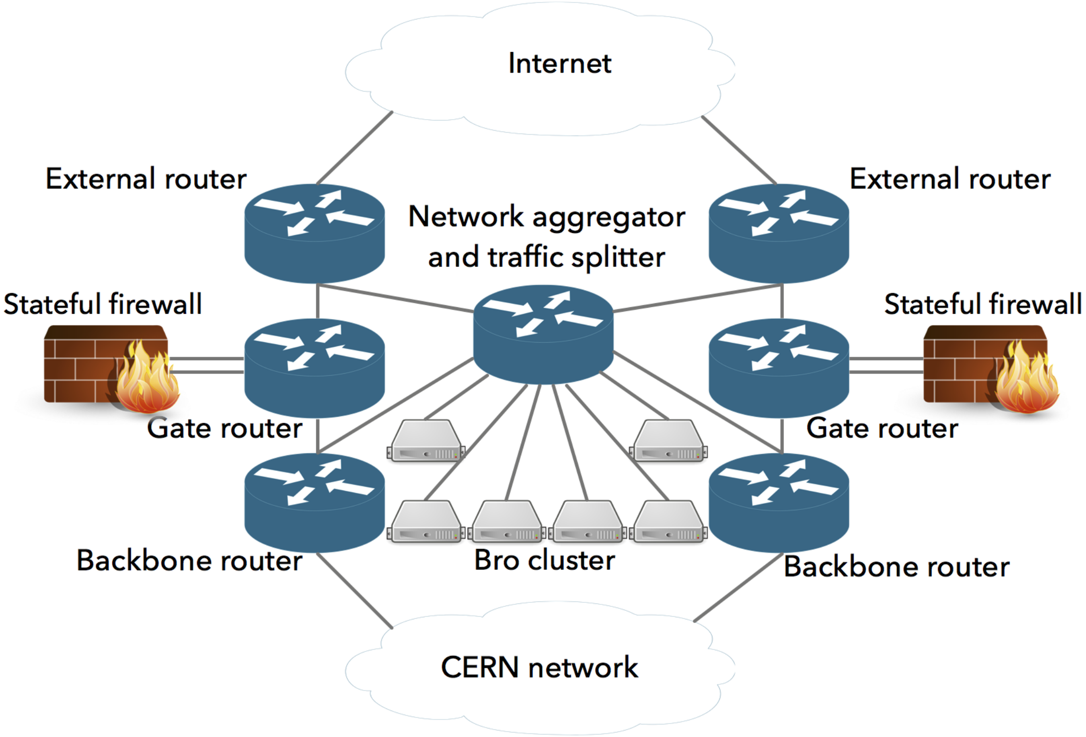

Figure 3: Network traffic aggregation and splitting at CERN.

The above networking setup is performing a splitting of the network traffic. A second stage splitting is then being performed inside the Bro nodes through the use of PF_RING. Bro packages are built at CERN with PF_RING support as the upstream packages are lacking built in support for it. Bro built packages are shared within the working group and with the rest of the HEP community [4].

A cluster comprised of a total of 16 Bro IDS nodes is being used, each node being capable of monitoring up to 10 Gb/s of network traffic. The specifications of the Bro nodes are:

- Intel S2600KP Kennedy Pass compute nodes (4 system units / chassis)
- 2x Intel Xeon E5-2630 v3 CPUs @ 2.40GHz (8 physical cores / CPU, 32 virtual cores in total / system)
- 128 GB of DDR4 memory running at 2133 MHz (8x 16 GB memory DIMMs)
- 2x 960 GB SSDs
- Intel X540-AT2 10-Gigabit on-board NICs and Intel 82599ES 10-Gigabit SFI/SFP+ add-on NICs used for monitoring the networking traffic

3.5 Threat Intelligence: MISP

The threat intelligence platform of choice for CERN is MISP. It allows quick and efficient sharing of threat intelligence data with a multitude of peer organizations. CERN is operating a total of three MISP instances:

- Main CERN production instance, synchronised with a number of external MISP instances spanning both academia and industry. As of March 2017 this instance was containing more than 460 000 IoCs spanning more than 5 400 events.

- WLCG production instance, set up for shared threat intelligence inside the working group and used as a central instance for the community. All the events reaching the main CERN MISP instance that CERN has the permission to redistribute are automatically synchronised to the WLCG instance. As of March 2017 this instance was containing more than 150 000 IoCs spanning more than 1 700 events.

- Development instance used for development work and testing of upgrades to newer MISP releases.

CERN has been actively contributing to the MISP project. Among the MISP developments done at CERN and integrated upstream we can mention:

- Export from MISP into Bro's intelligence framework native format

- Single Sign On authentication support

- Puppet module for deploying, configuring and updating MISP

3.6 Integration between MISP and Bro

Every 15 minutes a full export from MISP to Bro of the attributes (IoCs) for all the events published in the past 30 days is performed. The export is done in the native format used by the Bro intelligence framework while providing the context of the IoC, including:

- UUID of the MISP event

- The name of the organization which produced the event

- The MISP instance from where the event was exported

- Description of the event

- Description of the attribute (IoC)

The entries in Bro's intelligence framework are set to expire every 20 minutes. Given that the export is performed every 15 minutes, this allows for the natural expiration of IoCs exported to Bro.

4. Working group SOC development

Although using Metron as a general reference, the working group has followed closely the work on the CERN SOC. In that context, the following components were selected as the initial base from which to start development:

- Threat intelligence
- Intrusion Detection System

4.1 Threat Intelligence: MISP

Threat intelligence allows signatures of security events (Indicators of Compromise, or IoCs [5]) to be shared between participating sites. Within a given community, this allows the details of a security event at one site to be shared with others, providing the needed threat intelligence required to increase the protection of participating sites. Critically, this is an environment in which collaboration is key; sites of all types and sizes can contribute to the security of all sites, by contributing intelligence or by being able to act on it. A fundamental part of this work is the creation of trust frameworks. The tool under consideration by the working group is MISP [6].

MISP, or Malware Information Sharing Platform, is a web based application which provides a platform that enables subtle and flexible sharing of threat intelligence. Given the potential sensitivity of information being shared, MISP allows for information to be tagged appropriately, and sharing between instances can be regulated accordingly. Sharing can take place between organisations, within organisations and externally. In addition, an API can be used to pull data from a given MISP instance. As such, one may define a number of testing and deployment stages:

- Test installation
- Data sharing
- Data extraction via API

At this stage the group is at the phase of test installation and data sharing.

4.1.1 Current status

WLCG MISP instance

- Deployed at CERN
- Accessible through federated identity (eduGAIN [7])
- Federated Identity providers must have the SIRTFI [8] flag set

STFC/RAL

Two test instances were deployed at STFC/RAL within the Scientific Computing Dept. to investigate information exchange and sharing configurations. Further integration options are being evaluated.

Glasgow MISP instance

Glasgow's MISP instance was transitioned from a dedicated server to being deployed on a VM. IoC data is synced to this instance using a set of CERN credentials (pull only), which is then accessible locally to the Glasgow campus security team.

In addition, access to the WLCG MISP instance using the University of Glasgow Identity Provider has now been demonstrated after the appropriate SIRTFI flag was set.

4.2 Intrusion Detection Systems

Intrusion detection systems (IDS) can act both to detect anomalous behaviour on a host network, as well as acting on supplied intelligence. They act by monitoring (typically network) traffic through a well defined point in the site infrastructure. In particular, in this context they act as a source of data to be ingested into the SOC.

In that light we introduce the Bro [9] Open Source Network IDS (Intrusion Detection System) that is in broad use both across academia and industry. There are a number of large scale Bro IDS deployments in the US, one such example being at Berkeley Lab [10]. While this IDS is starting to gain more traction in Europe, it's one area that could benefit from more investigation, which is exactly what the current working group is aiming at.

4.2.1 Current status

Brunel

Brunel is currently investigating use of Bro with CISCO Nexus switches/Netflow.

Durham

Durham has been working through deployment issues, including data capture and volume, where they saw a reduction in capture loss when PF_RING was deployed. Currently their Bro instance is operated periodically due to degradation of network performance while Bro is running; future plans include the investigation of other networking equipment. Other details of their deployment include:

- Host: Lenovo nx360 - 40 HT cores
- Network: HP 5412zl mirroring 10gbit uplink (4gbit rate limited)
- 10 GB/day (raw)
- PF_RING used for traffic splitting inside the system

Glasgow

Glasgow originally deployed Bro on a 16 core server, mirroring their 10 Gb/s WAN uplink. After this showed a significant level of packets dropped at the Bro level [11] (non-uniform and peaking to 20-30%), a new deployment was carried out on a 64 core AMD Interlagos worker node. With this deployment, packet drop fell to an average of 1-3% with spikes to 10%. Specific details of the Glasgow installations include:

- Original host: R410, 16 HT cores
- Second host: C6145, 64 AMD Interlagos cores
- PF_RING used for traffic splitting inside the system
- Second deployment used CERN Bro RPMs
- 35 GB / day (raw) logs
- Network: Extreme x670, mirroring main 10Gb/s WAN uplink

In both cases, capture loss (incomplete packet capture) was considerable - in the first deployment it was seen to be typically 90%, while in the later deployment it was typically 70%. The source of this capture loss is an area of active development.

Lancaster

Lancaster is considering deployment with CERN Bro RPMs.

5. Future work

In the immediate future, the focus of the working group will be on the deployment strategies for Bro; given the close ties with the specific network topology of a given site, this has the most complexity in installation between sites. As such we will continue to deploy Bro at a number of sites, with the goal of covering different types of site.

In terms of MISP specific work, the next stage is to continue developing experience in the context of intelligence sharing and, in specific, testing accessing data via the MISP API.

Additionally, and at this point bringing the work on MISP and Bro together, we plan to investigate outside of CERN the integration of MISP and Bro, automatically importing data from MISP into Bro. There is a specific plan to investigate this at Glasgow.

Further, we will examine additional components to add to the SOC stack. It is likely that this will include the use of Logstash/Elasticsearch/Kibana to allow for advanced analysis and visualisation of the data. We will report on this progress in future work.

Finally, we will track the development of the Metron project. While this is more complex than fully appropriate for the group at this stage, it forms a useful milestone and a fruitful source of data as this work continues.

6. Conclusion

We have presented in this paper the current status of the work of the WLCG Security Operations Centers Working Group. This work is at a key stage where further involvement of other sites can greatly enhance the work, in particular in documenting experience with the Bro IDS as well as the integration of Bro and MISP.

References

[1] http://metron.incubator.apache.org

[2] http://github.com/opensoc/opensoc

[3] https://github.com/CERN-CERT/activity_klog

[4] http://linuxsoft.cern.ch/internal/repos/sec7-external/

[5] https://en.wikipedia.org/wiki/Indicator_of_compromise

[6] http://misp-project.org

[7] https://www.geant.org/Services/Trust_identity_and_security/eduGAIN

[8] https://refeds.org/sirtfi

[9] http://bro.org

[10] http://commons.lbl.gov/download/attachments/120063098/100GIntrusionDetection.pdf

[11] https://www.bro.org/documentation/faq.html#how-can-i-reduce-the-amount-of-captureloss-or-dropped-packets-notices