Collaborative Operational Security: The future of Cybersecurity for Research and Education

David Crooks[1] *James* Acris[1], *Liam* Atherton[1], *Paul* Clark[2], *Pau* Cutrina[3], *David* Jordan[4], *Shawn* McKee[5], and *Liviu* Vâlsan[2]
for the WLCG Security Operations Centers Working Group

[1]RAL, UKRI STFC, UK
[2]University of Durham
[3]European Organization for Nuclear Research (CERN), Geneva, Switzerland
[4]University of Chicago
[5]Physics Department, University of Michigan, Ann Arbor, MI, USA

Abstract. No single organisation has the resources to defend its services alone against most modern malicious actors and so we must protect ourselves as a community. In the face of determined and well-resourced attackers, we must actively collaborate in this effort across HEP and more broadly across Research and Education (R&E).
Parallel efforts are necessary to respond appropriately to this requirement. We must share threat intelligence about ongoing cybersecurity incidents with our trusted partners and deploy the fine-grained security network monitoring necessary to make active use of this intelligence. We must also engage with senior management in our organizations to ensure that we work alongside any broader organisational cybersecurity development programs.
We report on progress of the Security Operations Center (SOC) Working Group, established by the WLCG but with membership encompassing the R&E sector. The goal of the Working Group is to develop reference designs for SOC deployments and empower R&E organisations to collect, leverage, and act upon targeted, contextualized, actionable threat intelligence. This report will include recent SOC deployment activities at sites with network connectivity in excess of 100Gb/s, as well as new technology designs. An important development, which is likely to form a key part of the WLCG security strategy, is the potential use of passive DNS logs to allow sites without fine-grained network monitoring to benefit from the threat intelligence available to our community.
We also report on higher-level progress in engaging with the broader community to establish common approaches to this vital area of cybersecurity.

1 Introduction

As reported previously [1], the threat landscape faced by the research and education sector, including HEP, includes the possibility of attack from a variety of vectors, including phishing leading to ransomware. Since our last article on this topic, this threat is now acute, having grown in severity in recent years. There are several examples in the public press of attacks against research and education organisations that have led to extended downtimes and, in at

least one instance, the closure of the organisation. The impact of these site closures includes significant costs, both financial and reputational.

In this context, the work to improve the collaboration between R&E organisations to share threat intelligence and methods to actively use this intelligence has increased in importance. In this paper, we discuss progress in the Security Operations Centers Working Group (SOC WG).

2 Core principle

The core principle of this work is that we must share ongoing information about security incidents within our community in a way that can be systematically integrated into our organisational monitoring systems. The volume of potential attacks is such that manual intervention at the detection stage, for example, by manual searching of logs, is impractical.

Thus, we must have methods to automatically share information, or threat intelligence, which can then be automatically ingested into an integrated set of monitoring, storage, visualisation, and processing tools either on premise at an organisation or distributed across a region or infrastructure. This integrated set of tools, along with the attendant processes and staffing, is what we term a Security Operations Center for the purpose of this work.

3 Reference Design

The key elements of the original SOC WG reference design were the presence of a threat intelligence component – for which the MISP [2] platform is a critical part based on its wide use – along with the use of a fine-grained network monitoring system such as Zeek [3] [1]. The output of this data source was then ingested via a logstash pipeline to Elasticsearch, visualized with Kibana. The final stage of this design required a means of alerting.

Following the recent SOC WG Hackathon [4], the working group has developed a draft version of a second version of this reference design which allows for different models to be layered over an expanded set of base elements shown in figure 1. This should be read in conjunction with a set of questions designed to inform the choices for the tooling and process for each element, as shown in Table 1.

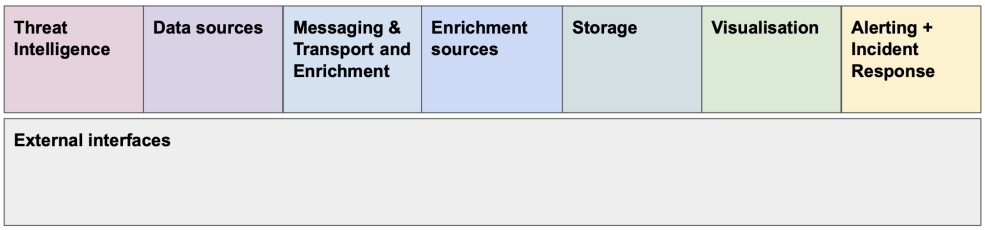

Figure 1: Draft second version of SOC reference design elements

A key new element of this updated reference design is the identification of *External interfaces*. For each element, how would this interface with other capabilities as part of an interoperable network?

[1] The potential for the use of net- or s-flow was included as a fallback or large scale solution, but this is less preferable as it does not provide forensic level logging through the typical use of sampling

SOC Element	Description
Threat Intelligence	By what means is threat intelligence shared, both strategic, and real-time and actionable
Data Sources	Which security data sources are available to the SOC, including network, endpoint, and logging sources
Messaging + Transport and Enrichment	How is data gathered from the data sources and made available to the rest of the SOC?
Enrichment sources	How is data enriched and presented to the storage platform?
Storage	How is the processed and enriched data stored?
Visualisation	How is the processed and enriched data visualised using dashboards or other mediums?
Alerting + Incident Response	How are actionable alerts raised based on processed and enriched data, correlated with threat intelligence; How is this made available to responders, and further integrated into new or existing incident management systems

Table 1: Questions used to inform choices for each of the base SOC elements

3.1 External interfaces

Looking across the organisational landscape of research and education, different institutions will typically use different tools and vendors to deploy a SOC capability. Although the SOC WG provides reference designs, it is clear that we cannot assert that all organisations must follow this design using the tools we recommend. Instead, as we look at the distributed environment, what is most important are the ways in which our SOC capabilities *interoperate* – at the highest level, following our core principle, what we need is that intelligence be shared and actively used rather than that all organizations deploy their tooling in exactly the same way. This would be an intractable challenge given the wide range of demands on organisations that will lead to different choices being made.

What is important at this level, therefore, is that we work towards an understanding of the relevant interfaces that will aid in collaboration between organisations towards a common defence.

In Section 5 we will explore a number of SOC models, based on the resources and experience available to a given entity wishing to deploy an SOC capability. We break these down into *Lightweight*, *Essential* and *Maturing*. In each case, we look at considerations of which tools may be most appropriate to organisations with different levels of resourcing and maturity.

Before we explore different SOC models, however, we introduce a new SOC component that would allow the use of DNS logs to provide a lightweight approach to a SOC deployment.

4 pDNSSOC

pDNSSOC is an open-source tool designed to transmit, receive, and analyze DNS logs. It can provide timely alerts about suspicious domains and IP addresses by combining DNS data with indicators of compromise in MISP. In addition to that, when integrated into a recursive DNS server, pDNSSOC transforms DNS traffic into a more complete dnstap format.

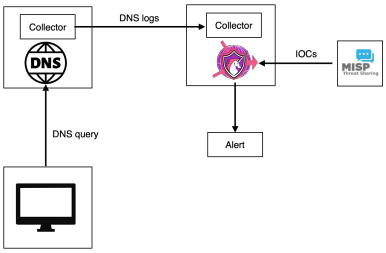

Figure 2: pDNSSOC Workflow

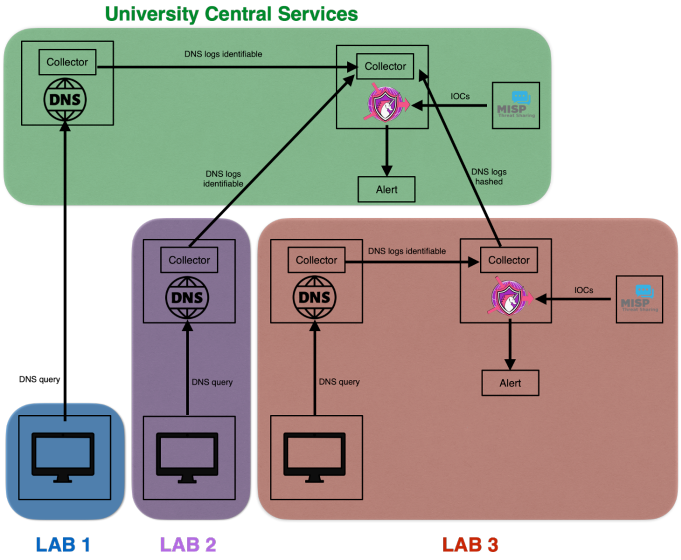

Figure 3: Possible pDNSSOC deployment configuration

The objectives of this project are clear. First, our aim is to provide security tools to all organisations, regardless of their expertise, maturity or financial capability. Second, we aspire to emphasise the importance of shared intelligence, establishing it as the main resource for detecting and addressing incidents. Furthermore, it is important to encourage strong and impactful collaborations between institutions in similar sectors. Finally, we want to expand the visibility of central security teams across affiliated resources, organisations, and communities.

5 SOC Models

Having established the new pDNSSOC component, we may now look at potential SOC models using the new draught SOC reference design. As mentioned above, we would currently consider *Lightweight*, *Essential* and *Sustained* models. Of these, the Essential model is most closely related to the original reference design, although with some additions in the Alerting and Incident Response element

SOC Element	Description	External Interfaces
Threat Intelligence	Email	
Data Sources	(p)DNS agent; DNS logs; Service logs	remote pDNSSOC
Messaging & Transport and Enrichment		
Enrichment sources		
Storage		
Visualisation		
Alerting + Incident Response	Staff availability to field alerts	remote pDNSSOC

Table 2: Proposed tooling and process for the Lightweight SOC Model

5.1 Lightweight SOC

The context of the lightweight SOC model is where an organisation or facility does not have a mature cybersecurity programme and does not have the internal resource to develop such a programme in the short term; this may of course change with time.

In this case, the simplest components are indicated in table 2.

5.2 Essential SOC

The Essential SOC model is intended as the minimum viable product for an organisation that is beginning a long-term SOC deployment programme; the tooling and processes proposed for this model can be seen in Table 3.

5.3 Sustained SOC programme

The extension of the Essential SOC model is a sustained development of the Security Operations Centre, following the community in terms of the most important data sources and enrichment sources.

6 Status updates

6.1 Durham

Security at Durham continues to be an important aspect of daily site operations; however, due to ever changing hardware specifications, currently finds itself in a difficult position. Currently the site deploys a single Zeek node monitoring a mirrored 40Gbps bond, this is updated from the original specification of a 20Gbps bonded link. By doubling data traffic without increased hardware upgrades, this is ultimately beginning to cause issues. Plans to rebuild this service to permit a 100Gbps upgrade are in place and will be rolled out in the near future. Likewise, the issues surrounding the Zeek system have had a negative effect in terms of security monitoring due to data inaccuracies.

A full rebuild is planned for 2024 with designs and specifications currently in the planning phase, these include additional integrations with system components throughout Durham to

SOC Element	Description	External Interfaces
Threat Intelligence	Access to MISP with actionable threat intelligence or appropriately synchronised local MISP instance	MISP; other threat Intel formats
Data Sources	Zeek (Corelight) and Curated/ Prioritised logs	
Messaging & Transport and Enrichment	Necessary data source plugins / agents (e.g. Filebeat); Kafka (based on criteria to be specified); Logstash	Kafka
Enrichment sources	GeoIP; DNS enrichment (reverse lookups and FQDN/IP); CRIC (WLCG); RIR info (ASN/WHOIS/etc)	pDNS; Vendors
Storage	Opensearch	
Visualisation	Opensearch Dashboards + Grafana	
Alerting + Incident Response	Aggregation Scripts; Zeek Intel framework; Elastalert; FIR	Ticketing integration; email

Table 3: Proposed tooling and process for the Essential SOC Model

aid in the creation and dissemination of usable threat intelligence for use in both a grid and the local environment. Moving forward, this will include a rebuild/migration of the current elastic cluster to Opensearch along with the addition of components that focus on data enrichment to help provide actionable outputs.

6.2 MWT2 - University of Chicago

MWT2 at the University of Chicago is building on its increasing push for IT security. We have deployed two Zeek nodes. One for monitoring the 2x100Gb active link and one for the 2x100Gb backup link into the network. These are placed between the MWT2 edge switches and the UChicago network, using optical tap cassettes to route a copy of the network traffic to Zeek. MWT2 currently has no plans to upgrade the network at the University of Chicago or to set up Zeek at its other two data centres in Indianapolis or at the University of Illinois Urbana-Champaign.

MWT2 has a MISP instance locally on a VM stack within the MWT2 network. This pulls threat intelligence from the central CERN instance. Zeek is set up to notify and alert based on intelligence it pulls from the MWT2 MISP and what traffic it sees coming through the network. MWT2 does have an Elasticsearch cluster but is currently not integrating it, or adjacent software, e.g. Elastiflow/Elastalert, with Zeek and MISP.

6.3 CERN

CERN has been operating a mature full-scale SOC for more than 5 years. Recent evolutions include the expansion of data sources, the migration from ElasticSearch to OpenSearch, and the redesign of incident response tools.

In terms of data ingestion a number of additional sources of data were added, including:

- Logs from the next-generation firewall, including traffic logs and threat logs.

- The migration of traceability logs from the kernel modules execlog and netlog to the Linux audit framework-based traceability logging via the use of Auditbeat and Packetbeat.
- Early investigations of ingesting cloud logs, specifically from Microsoft Azure and Google Clould Platform.

With the change in Elasticsearch licencing, CERN has taken the decision to migrate to Opensearch. The CERN SOC is currently making use of three Opensearch clusters, storing a combined 15 billion log entries / day, with a data volume of about 5 TB / day and a retention period of 60 days (HDFS storage comes with a longer retention period).

Lastly, incident response tooling has been updated with additional endpoint malware detection capabilities (Threatray) and remote forensic capabilities (Velociraptor).

A block diagram of the CERN SOC can be found in Figure 4.

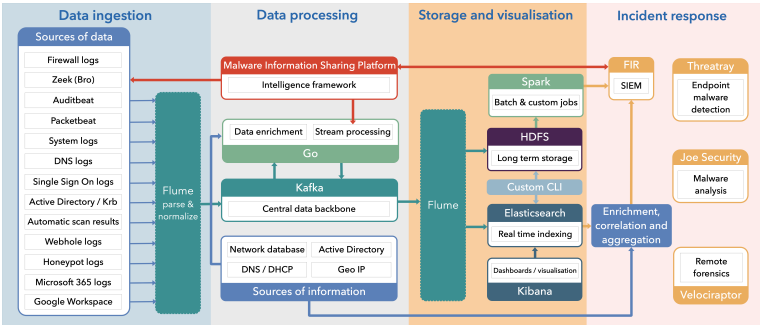

Figure 4: CERN SOC block diagram

6.4 STFC SOC Phase One

The STFC SOC has deployed and configured all hardware, including optical taps, on all links into the RAL site. This consists of 2x100Gb/s links for JANET and 2x100Gb/s links for LHCOPN, resulting in a total throughput of 400Gb/s. We have set up two zeek nodes ingesting data from the optical taps to monitor the links. One is in production, whilst the other is still being used for development. We are in the process of replacing our MISP instance with a more robust and easy to maintain containerised deployment, which has been developed by JISC, having also tested tools for Zeek to ingest threat intelligence from MISP. The data will be piped from Zeek into a Kafka cluster, and from there we intend to use Logstash to ship the data to an OpenSearch cluster for longer-term storage, retrospective analysis, and further monitoring. Our OpenSearch cluster is in production, though the pipeline is still under development. An important piece of work for us which underpins the SOC was to create a base set of templates in our configuration management system which prioritises the principle of least privilege, disabling root login to machines by default to minimise the risk of privilege escalation, and tighter access control than in other areas of the department.

7 Operational Security

The work of the SOC working group is focused on the technology element of a SOC deployment, but coupled to this work must be an accompanying work on the operational side. We focus here on two aspects of the operational use of threat intelligence, the EGI CSIRT and SAFER.

7.1 EGI CSIRT

Historically, the incident response team of EGI CSIRT, IRTF, has distributed indicators of compromise (IoCs) associated with a given incident through email broadcast to the security contact community within the EGI.

Current technical developments allow the IRTF to develop MISP events to share within the WLCG community; work is ongoing to incorporate this into the team's response procedures.

7.2 SAFER

SAFER is an operational security trust group focused on fighting computer misuse and defending the academic, research and education mission as a global community [5].

8 Conclusions

The development of the second version of the SOC WG reference design allows for increased flexibility in the deployment of SOC tools and processes at sites with different levels of maturity in cybersecurity and resource availability.

Work must now continue in our community to deploy these capabilities, coupled with an ongoing effort to integrate these tools with our operational security teams to provide sources of accurate and timely threat intelligence to enable our best defence against our growing cybersecurity risk.

9 Acknowledgements

We acknowledge our collaborations with the CERN IT, WLCG and LHCONE/LHCOPN communities who also participated in this effort.

References

[1] D. Crooks, L. Vâlsan, K. Mohammad, S. McKee, P. Clark, A. Boutcher, A. Padée, M. Wójcik, H. Giemza and B. Kreukniet, Operational security, threat intelligence & distributed computing: the WLCG Security Operations Center Working Group. EPJ Web Conf., 2019

[2] https://www.misp-project.org

[3] https://zeek.org

[4] https://indico.cern.ch/event/1268239/

[5] https://www.safer-trust.org/