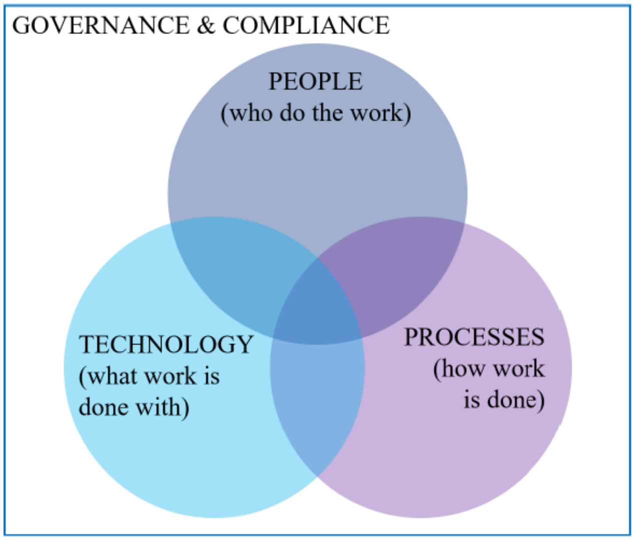
# Security Operations Center: A Systematic Study and Open Challenges

Many attacks go undetected for a surprisingly long time. The mean time to detect an incident was 196 days in 2018, and it took another 69 days on average to contain the breach. This detection time demonstrates how ineffective companies are at detecting and mitigating cyber-attacks.

A SOC is an organizational unit operating at the heart of all security operations. It is usually not seen as a single entity or system but rather as a complex structure to manage and enhance an organization's overall security posture. Its function is to detect, analyze, and respond to cybersecurity threats and incidents employing people, processes, and technology.


GOVERNANCE & COMPLIANCE
PEOPLE (who do the work)
TECHNOLOGY (what work is done with)
PROCESSES (how work is done)

**Triage Specialist: Tier 1 analysts** are mainly responsible for collecting raw data as well as reviewing alarms and alerts. They need to confirm, determine, or adjust the criticality of alerts and enrich them with relevant data. For every alert, the triage specialist has to identify whether it is justified or a false positive. If occurring problems cannot be solved at this level, they are escalated to tier 2 analysts.

**Incident Responder: Tier 2 level analysts** review the more critical security incidents escalated by triage specialists and do a more in-depth assessment using threat intelligence (Indicators of Compromise, updated rules, etc.). They need to understand the scope of an attack and be aware of the affected systems. If a tier 2 analyst faces major issues with identifying or mitigating an attack, the incident is escalated to tier 3.

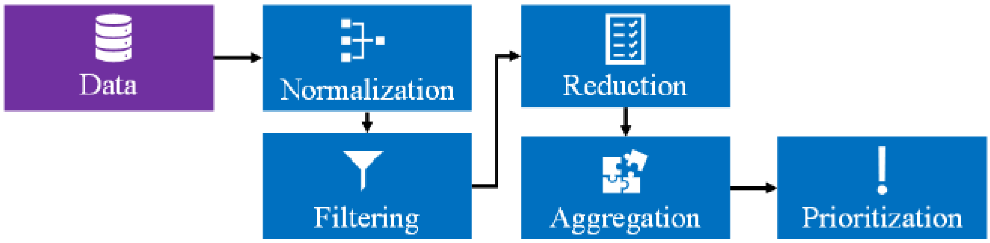
**Threat Hunter: Tier 3 analysts** are the most experienced workforce in a SOC. They handle major incidents escalated to them from the incident responders. They also perform or at least supervise vulnerability assessments and penetration tests to identify possible attack vectors. Their most important responsibility is to proactively identify possible threats, security gaps, and vulnerabilities that might be unknown.

**SOC managers** supervise the security operations team. They provide technical guidance if needed, but most importantly, they are in charge of adequately managing the team. This includes hiring, training, and evaluating team members, creating processes, assessing incident reports, and developing as well as implementing necessary crisis communication plans.

Besides technical skills, soft skills are becoming more and more important. Desired skills include communication skills, continuous learning abilities, analytical mindset, ability to perform under stress, commitment, teamwork, curiosity, and practical organizational skills.

Several methods to counteract staff burnout and increase job satisfaction can be determined:

- Increase Automation: Increasing automation helps decrease the amount of mundane and boring tasks.
- The more skills employees master, the more likely they are to be empowered. This empowerment enables employees to do their job efficiently and increases their morale.



Data collection process steps:

- **Normalization**: It is vital to translate the heterogeneous data formats into a uniform representation to conduct further processing. It is also essential to change all time data to one standard time zone and format.

- **Synchronization** helps avoid confusion in the timeline of the security events and reduces the likelihood that erroneous conclusions are made on inconsistently measured network activity. In literature, normalization is often referred to as log parsing or pre-processing.

- **Filtering**: Since systems typically generate enormous amounts of data, it is essential to filter for data elements that are likely to contain important information from a security perspective.

- **Reduction**: Reduction is like filtering, with the difference that individual, unimportant data fields are sorted out to reduce the amount of data.

- **Aggregation**: Similar events are combined into one single data element. For example, three log entries, which indicate a log attempt to a host, could be aggregated to one single log, which states the type and number of login attempts.

- **Prioritization**: Each log data should be classified according to importance to facilitate further processing.

The sheer amount of data collected in previous steps can be overwhelming, even for seasoned security practitioners and researchers. Turning this data into useful information is done through data analysis and is essentially a means to make sense of what is collected.

Incident Response Steps:

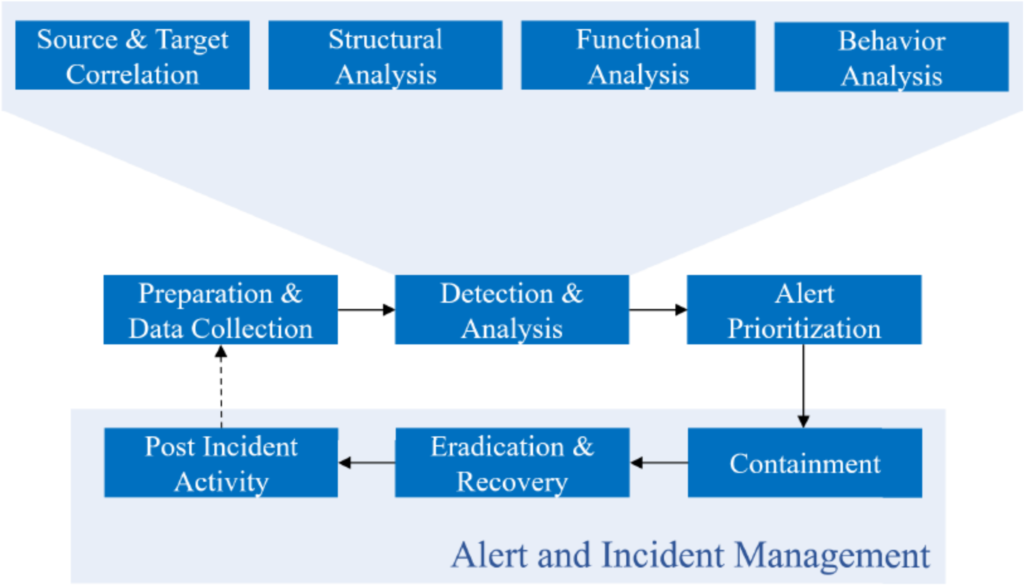
- **Detection**: Incidents are detected with the help of humans or by automatic procedures. Thereby, it must be decided if the collected data indicates a security incident.

- **Analysis**: Regarding the techniques used for analysis, one can distinguish between source and target correlation, structural analysis, functional analysis, and behavior analysis.

- **Alert prioritization**, also known as triage, can be seen as a link to containment, eradication, and recovery. It serves two primary purposes. First, to ensure that the most severe incidents are treated with priority, and second, to ensure that incidents are distributed for further processing according to available resources.

- **Containment**, eradication, and recovery... this step aims to decide whether an incident is an unharmful event (e.g., during penetration testing), or a harmful event. In the case of a harmful incident, it is passed on to appropriate stakeholders to take further steps.

The functionalities of SOAR are mainly categorized into integration, orchestration and automation. Security orchestration is a prerequisite of security automation, which is the process of automatic detection. Therefore, SOAR integrates available information about security incidents (Cyber Threat Intelligence) to automatically take appropriate measures to limit the damage as quickly as possible.



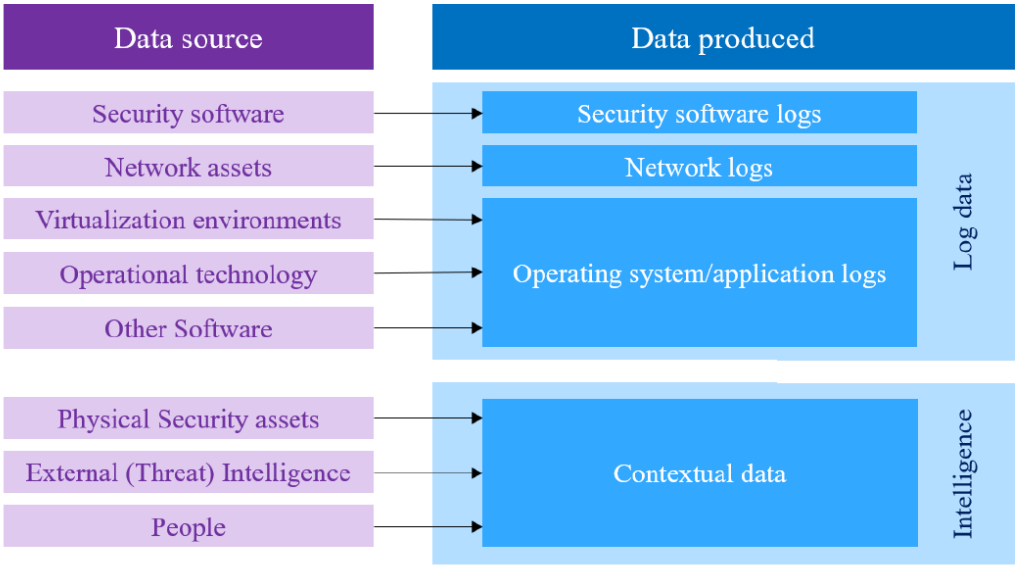
While new data sources are continuously being created, the most common sources, its classification, and corresponding examples are:

- Security software: SIEM systems, intrusion detection/ prevention systems, firewalls, anti-virus software, vulnerability scanners
- Network assets: Switches, routers, servers, hosts, proxies.
- Virtualization environments: Hypervisor, virtual machine introspection, cloud environments

- Operational technology: Sensors, actuators, PLCs
- Other Software: Databases, identity and access management, mailserver, operating systems
- Physical security assets: Security cameras, access control
- People: Employees (Human-as-a-Security-Sensor), external users.

Each of these data sources can deliver a vast amount of information, of which not all is relevant. Capturing everything may help in spotting malicious activity, but it can also negatively impact system performance. Conversely, if fewer data sources are used to collect data, an attack might go undetected.

Thus, finding the right balance between capturing too much and capturing too little data is essential when designing a SOC's technological capabilities. However, as a rule of thumb, it is generally better to capture data from as many sources as possible (under performance constraints) and then rely on well established data normalization, correlation, and analysis mechanisms.



Attack detection is performed either automatically or manually. Manual detection is the detection of an incident through an internal or external person. Manual detection is necessary, because not all attacks can be detected through technology, especially when it comes to advanced attacks.

Anomaly-based or behavior-based methodologies use the system's normal behavior as a foundation and try to detect deviations. Signature-based or also knowledge-based methods use accumulated knowledge of attacks and is very useful to detect known attacks or exploitation of known system vulnerabilities. Therefore, it is important to regularly update the knowledge base. Specification-based methodologies focus on detecting incidents based on predefined profiles or protocols. Hybrid methodologies use a mixture of the three described detection methodologies.

Concerning detection approaches, statistics-based detection is one of the oldest methods used for intrusion detection and uses statistical properties and statistical tests like mean, median or variance, to detect deviation between the normal behavior and observed behavior. Threshold metrics, hidden Markov models and multivariate models are examples of statistical based detection approaches. Pattern-based and Rule- based approaches use either predefined patterns, learned patterns or rules for detection.

Heuristic-based approaches are inspired by biological concepts as for example artificial neural networks. State-based approaches try to infer the behavior of attacks within the network for example by utilizing infinite state machines.

As there is no abstract, high-level understanding of a SOC's processes, many researchers focus on trying to improve technologies that might be useful with no clear understanding of which specific process or task of a SOC needs improvement. Also, having a clear understanding of a SOC's processes, tasks, and interfaces requires the integration with other business processes.

We see three major challenges for SOCs resulting from the increased complexity of the IT and OT environment in a company: First, the infrastructure is becoming more complicated and intertwined, making it difficult to maintain situational awareness and a cohesive overview. Managers and analysts have poor visibility into the network because they cannot keep track of all the devices in the network.

Second, the data captured from the infrastructure is as heterogeneous as its sources, making it hard to process, analyze, understand, and link. It also impedes the discovery of whether an event is part of a bigger attack.

Third, having more data sources increases the overall number of events and, in many cases, the number of false-positive alerts. It is often mentioned that there is too much (useless) data in general, and too many (false positive) alerts.