DiSIEM: Diversity-enhancements for Security Information and Event Management

(H2020 project 700692)

Very Short Project Overview

Contact: Alysson Bessani (project coordinator) – anbessani@ciencias.ulisboa.pt

Period: September 2016 – August 2019.

List of participants

Participant No	Participant organisation name	Organisation Short Name	Country
1 (Coordinator)	Fundação Faculdade de Ciências da Universidade de Lisboa	FFCUL	PT
2	City University London	CITY	UK
3	Energias de Portugal SA	EDP	PT
4	Amadeus SAS	Amadeus	SP
5	DigitalMR	DigitalMR	UK
6	Fraunhofer Institute	FHG	DE
7	Atos Spain SA	Atos	SP

Project Key Facts

- Although a **fundamental tool in modern Security Operations Centres**, current **SIEMs have many limitations** on the methods and means they use to collect events, store data and report information.
- The cornerstone of the DiSIEM project is the use of **scalable information extraction and machine learning algorithms** and tools to extract information from multiple big data sources (sensors in the monitored infrastructures, **open-source intelligence**, social networks, security newsfeeds, advisory organizations, etc.) and feed SIEMs with it for threat prediction and enhanced risk assessment, aided by probabilistic methods and advanced visualisation tools.
- DiSIEM wants also to equip existing SIEM systems with **the capability of evaluating diverse configurations of monitoring and protection devices**, novel **application-based misuse detection** and **secure cloud-backed long-term archival of selected events**.
- The **DiSIEM components can be applied to any existing SIEM** that supports custom connectors and provides access to the event store.
- **DiSIEM components can be used either individually or together**, broadening the scope in which the project results can have impact.
- **DiSIEM components will be validated in production environment** for three large organisations: **an electricity utility company** (EDP), **a large travel services company** (Amadeus) and a **SIEM and security provider** (Atos).
- The **DiSIEM exploitation business model** considers components that will be supported by **partners offering services to SIEM operators** (DigitalMR, Atos), internally by **partners operating large SIEM** (Amadeus, EDP) and **startup initiatives created primarily from the research and development** partners (FFCUL, CITY, FHG).
- The project is organised in **6 technical work packages and 3 additional work packages** for ethics, dissemination, exploitation and management activities.
- The project will run for **three years**, starting in September 2016, with an overall budget of around **four million euros**.
- The consortium will be **assisted** and **advised** by an **advisory board** including representatives from public and private sectors.

Current Challenges to Security Management Systems

Organizations currently monitor and manage the security of their infrastructures by setting up Security Operation Centres (SOC) to make security-related decisions (e.g., which system is under attack, what has been compromised, where has an access breach occurred, how many attacks have happened in the last 12 hours). A SOC obtains an integrated view of the monitored infrastructure by employing a Security Information and Event Management (SIEM) system. These are complex systems that incorporate the functionality to collect logs and events from multiple sources, correlate these events together and then produce summarised measurements, data trends and different types of visualisations to help system administrators and other security professionals. Due to the nature of the functionality of these systems (the number of systems that feed events to them, the different types of events they need to correlate etc.) they are complex and costly to deploy and maintain.

The SIEM market is a growing one. According to a recent (July 2015) Gartner report (Magic Quadrant for Security Information and Event Management), in 2014 the SIEM market grew from $1.5 billion to approximately $1.69 billion, achieving a growth rate of about 14%. There are many high quality products from large IT vendors. Examples are IBM QRadar,[1] HP ArcSight,[2] Splunk,[3] LogRhythm[4] and AlienVault OSSIM.[5] Overall, the Gartner report identified two main drivers for such growth: threat management and compliance. The spectrum of new attacks (with hundreds of novel kinds of malware each month, including the ones related with advanced persistent threats) and the complexity of the IT infrastructures require a well-structured and integrated monitoring of security events. Additionally, many industries have strict requirements for compliance,[6] especially when it comes to log management, which often mandates the need for integrated log management functionality, as provided by SIEM systems.

Despite their widespread use and the impressive market growth, current SIEMs still have many limitations:

1. The **threat intelligence capacity of SIEMs is still in its infancy**. Consequently, the systems are unable to automatically recognize novel threats that may affect (whole, or parts of) the monitored infrastructure, requiring considerable human intervention to adapt and react to changes in the threat landscape. This happens despite the availability of rich and up-to-date security-related information sources on the Internet (e.g., social media, blogs, security newsfeeds), which current SIEMs are unable to use.
2. **Current systems can show any "low-level" data related with the received events, but they have little "intelligence" to process this data and extract high-level information**. These low-level data (e.g., number of failed logins in a server) are only accessible and meaningful to a limited subgroup of system admins, and are difficult to translate to high-level metrics for senior, C-level managers (such as executives and decision-makers who may need to make decisions on security expenditure, but may not necessarily be well versed in the technical details). This impacts, for instance, the capacity of SOC coordinators to justify the return on investment in security for an organization.
3. **Most data visualisation techniques in current SIEMs are rudimentary**. Advanced data visualisation in current SIEMs is still limited. This can seriously impact the ability of the SOCs to deal with incidents as and when they happen, in a timely manner.
4. The **event correlation capabilities of SIEMs are as good as the quality of the events fed to it**. Imprecise events and alarms generated by imperfect monitoring devices will be taken as correct by the SIEM and the uncertainties associated with these events are never communicated.
5. Due to storage and event processing constraints, **SIEMs are incapable of retaining the collected events for a long duration**. This limits their use in conducting forensic investigations in the long run, for example on advanced persistent threats, or other historical incidents.

The **Diversity-enhanced SIEM (DiSIEM)** project aims to address these limitations by complementing existing SIEMs with a set of components for accessing diverse data sources, feeding enhanced events to the SIEM and generating enhanced reports and metrics to better support the security operation centres.

[1] http://www-03.ibm.com/software/products/en/qradar-siem

[2] http://www8.hp.com/us/en/software-solutions/siem-security-information-event-management/

[3] http://www.splunk.com/

[4] https://www.logrhythm.com/

[5] https://www.alienvault.com/products/ossim

[6] http://www.sans.org/reading-room/whitepapers/auditing/successful-siem-log-management-strategies-audit-compliance-33528

DiSIEM Objectives

The **DiSIEM** project aims to address the limitations described above to improve SIEMs already deployed in production. Instead of proposing novel architectures for future SIEMs or modifications to existing ones, the project will address the aforementioned limitations by extending current systems, leveraging their built-in capacity for extension and customisation. The **core idea** of the project is to enhance existing SIEM systems with several diversity mechanisms, representing five main advances in the state of the art:

1. Integrate **diverse OSINT (Open Source Intelligence) data sources** available on the web, such as the NIST's National Vulnerability Databases, vulnerability and patch databases offered by vendors; threat intelligence data that organisations share with each other (e.g., Internet addresses, URLs and file reputation databases like SANS ISC, malware domains lists, VirusTotal, ThreatExpert, SpamHaus, OpenBL, EmergingThreats, etc.); security blogs and data streams from social networks (e.g., Twitter, Facebook, LinkedIn), collaborative platforms used in the Dark Web (e.g., Pastebin), search engines and online repositories (e.g., Google Hacking Database, Shodan, RIPE/ARIN, Whois), standards-based Indicators of Compromise or IOC's (e.g., STIX and OpenIOC), and many others. This data needs to be fetched, analysed, normalised and fused to identify relationships, trends and anomalies and hence help reacting to new vulnerabilities to the new infrastructure or even predict possible emerging threats against the infrastructure monitored by the SIEM.
2. Develop novel **probabilistic security models** and **risk-based metrics** to help security analysts to decide which infrastructure configurations offer better security guarantees and increase the capacity of SOCs to communicate the status of the organisation to C-level managers.
3. Design novel **visualisation methods to present the diverse live and archival data sets**, to better support the decision-making process by enabling the extraction of high-level security insight from the data which will be used by the security analysts working with SOCs that operate the SIEM.
4. In order to increase the value of the events fed to the system we will integrate **diverse, redundant and enhanced monitoring** capabilities to the SIEM ecosystem. The idea is to have enhanced sensors and protection tools built using a set of diverse tools. For example, by using three different intrusion detection systems to monitor the same critical part of the network, we can have a much higher confidence on the alarms generated by such systems. Implementing these kinds of mechanisms requires probabilistic modelling of diversity for security to define which combinations of tools are more effective and how much improvement can be expected. Likewise, we propose to deploy and integrate novel behavioural anomaly detectors for business-critical applications and thus improve the SIEM's visibility into the functional security status of these monitored applications.
5. Add support for long term archival of events in public cloud storage services. In order to satisfy the security requirements of such data (which contains a lot of sensitive information), **we will store such events in diverse cloud providers** (e.g., Amazon, Windows Azure, Google), employing techniques such as secret sharing and information dispersal.

These contributions would be materialized through a set of tools and components, in the form of plugins, that can be integrated into existing SIEM systems. For example, redundant diverse analysis and trends obtained through OSINT sources can be fed to the SIEM, while new visualization and analysis tools can be integrated by fetching data from the SIEM event database. The envisioned architecture of a SIEM implementation enhanced with DiSIEM contributions appears in Figure 1.2.

Expected Results

The main results emanating from this project will be the design and implementation of the several components illustrated in the red boxes in Figure 1.2:

- **Techniques and tools for analysing, evaluating and guiding the optimal deployment of diverse security mechanisms** in the managed infrastructure, including **multi-level risk-based metrics** (employed in all red boxes in the figure).
- An **OSINT-based security threat predictor** (the "OSINT Data Analysis and Fusion" box).
- A rich set of enhanced **interactive visualisations** for improving the quality of the decision support of security analysts operating a SIEM (the "Visualisation and Analysis Tools" box).

- A **framework for deploying diverse and redundant sensors** (part of the "Diversity-Enhanced Monitoring" box).
- A novel **application-based anomaly detector** for complementing other sensors and detect fraud in application servers ("Diversity-Enhanced Monitoring" box).
- Components that allow for long-term **event archival in diverse clouds** (the "Cloud-of-clouds Event Archival" box).

By choosing the extension approach instead of developing a new SIEM (or expecting vendors to change their systems to accommodate our enhancements), we expect to foster innovation much faster, and maximize the impact and business potential of the project results.

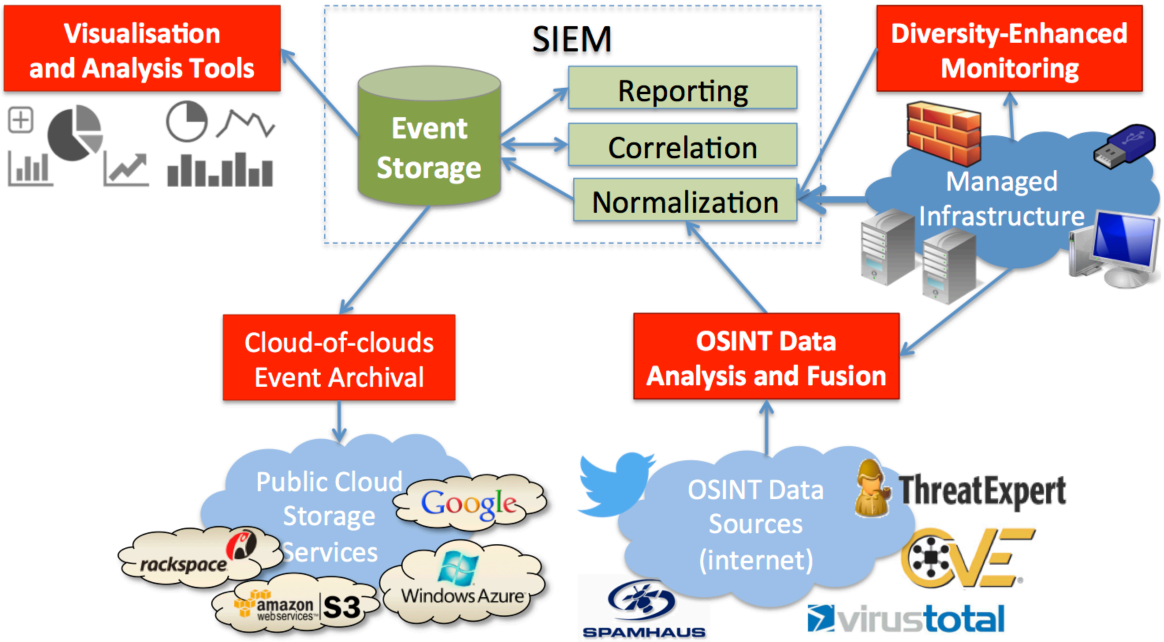
Visualisation and Analysis Tools
SIEM
Event Storage
Reporting
Correlation
Normalization
Diversity-Enhanced Monitoring
Managed Infrastructure
Cloud-of-clouds Event Archival
OSINT Data Analysis and Fusion
Public Cloud Storage Services
Google
rackspace
amazon webservices
S3
Windows Azure
OSINT Data Sources (internet)
ThreatExpert
CVE
SPAMHAUS
virustotal

Figure 1.2: DiSIEM architecture around an existing SIEM (red boxes represent our key contributions).