6

Cyber-Threat Intelligence from European-wide Sensor Network in SISSDEN

Edgardo Montes de Oca[1], Jart Armin[2] and Angelo Consoli[3]

[1]Montimage Eurl, 39 rue Bobillot, Paris, France
[2]CyberDefcon BV, Herengracht 282, 1016 BX Amsterdam, the Netherlands
[3]Eclexys Sagl, Via Dell Inglese 6, Riva San Vitale, Switzerland
E-mail: edgardo.montesdeoca@montimage.com; jart@cyberdefcon.com; angelo.consoli@eclexys.com

SISSDEN is a project aimed at improving the cyber security posture of EU entities and end users through development of situational awareness and sharing of actionable information. It builds on the experience of Shadowserver, a non-profit organization well known in the security community for its efforts in mitigation of botnet and malware propagation, free of charge victim notification services, and close collaboration with Law Enforcement Agencies (LEAs), national CERTs, and network providers. The core of SISSDEN is a worldwide sensor network which is deployed and operated by the project consortium. This passive threat data collection mechanism is complemented by behavioural analysis of malware and multiple external data sources. Actionable information produced by SISSDEN provides no-cost victim notification and remediation via organizations such as CERTs, ISPs, hosting providers and LEAs such as EC3. It will benefit SMEs and citizens which do not have the capability to resist threats alone, allowing them to participate in this global effort, and profit from the improved analysis and exchange of security intelligence, to effectively prevent and counter security breaches. The main goal of the project is the creation of multiple high-quality feeds of actionable security information that can be used for remediation purposes and for proactive tightening of computer defences. This is achieved through the development and deployment of a distributed sensor network

based on state-of-the-art honeypot and darknet technologies, the creation of a high-throughput data processing centre, and provisioning of in-depth analytics, metrics and reference datasets of the collected data.

6.1 Introduction

The primary data collection mechanism at the heart of the SISSDEN project[1] is a sensor network of honeypots and darknets. The sensor network is composed of VPS provider hosted nodes and nodes donated to the project by third-parties acting as endpoints. These VPS nodes/endpoints are not the actual honeypots themselves. Instead, they act as layer 2 tunnels to the SISSDEN datacenter. Attack/scan traffic to the VPS nodes is sent via these tunnels to corresponding VMs which run the actual honeypots themselves. The honeypots in the datacenter then respond to the attacks/scans with the IP addresses from the VPS nodes.

This approach allows for easier management of the honeypots themselves – instead of having to remotely manage (and maintain) honeypots at the VPS provider locations, all can be centrally managed in one datacenter instead.

Each sensor endpoint has multiple IPv4 addresses – one for management, the others for tunnelling to the real honeypots.

As of 14th of January 2019, SISSDEN has 226 operational nodes running, spread across 58 countries. A total of 953 IP address from 112 ASNs are used, covering 375/24 networks.

The following world map (Figure 6.1) shows the current snapshot of operational sensor IPs:

Nine different honeypot types are currently deployed. These are focused on observing different forms of attacks against SSH/telnet services, general or specialised web services, remote management protocols, databases, mail relays, ICS devices, etc, including exploits, scans, brute force attempts. Information about these attacks is disseminated to 95+ National CSIRTs and 4200+ network owners via Shadowserver's free daily remediation feeds. These are marked with source 'SISSDEN'. One can subscribe to SISSDEN feeds via the SISSDEN Customer Portal (https://portal.sissden.eu).

To capitalise on the tools and knowhow from the H2020 SISSDEN project and assure the sustainability of the results, innovative real-time Cyber Threat

[1]SISSDEN (Secure Information Sharing Sensor Delivery event Network) is an H2020 project. See https://cordis.europa.eu/project/rcn/202679_en.html and https://sissden.eu/ for more information.

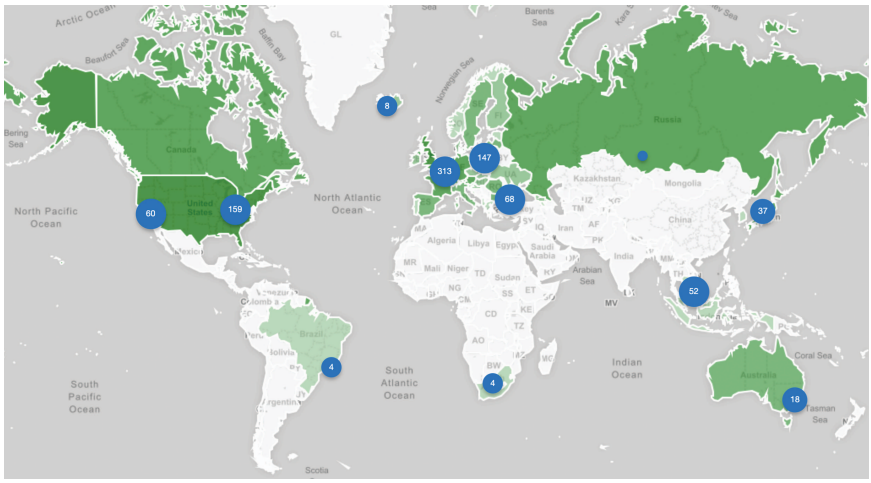

Figure 6.1 Map of deployed SISSDEN sensors.

Intelligence data for timely threat detection and prevention will be provided by a new start-up company called SISSDEN BV (https://sissden.com), launched by three SME partners (CyberDefcon, UK/The Netherlands, Montimage, France, and Eclexys, Switzerland).

6.2 SISSDEN Objectives and Results

6.2.1 Main SISSDEN Objectives

The main objectives of the SISSDEN project are:

- Create a large distributed sensor network. Over 100 passive sensors based on current and beyond state-of-the-art honeypot and darknet technologies are deployed in multiple organisations, including all 28 EU member states and 6 candidate countries, and are being used to observe malicious activities on an unprecedented scale, without intercepting any legitimate traffic.
- Advancements in attack detection. New types of honeypots, darknets and probes are deployed to detect, analyse and alert on types of attacks not widely detected today, such as reflective DDoS amplification or attacks against Internet of Things (IoT) devices, which are expected to increase significantly in the coming years as a range of new network-centric technologies are embraced by consumers and SMEs globally.

- Advancements in malware analysis and botnet tracking. The large sensor network is augmented by an innovative new generation of enhanced sandbox technologies designed for long running monitoring of malware specimen execution and behavioural clustering, to provide even more information on current threats.
- Improving the fight against botnets. Sensor and sandbox data collected is used for detailed studies of botnet infrastructures. Long-term observation of multiple families of current botnets will support anti-botnet research and law enforcement activities. Output will closely align with the existing European anti-botnet and anti-cybercrime strategies, as well as providing support to proven strong LEA partnerships, such as with Europol's European Cybercrime Center (EC3).
- Collect, store, analyse and reliably process Internet scale security data sets. The inherent challenges of building and continuously operating reliable data collection, storage, exchange, analysis and reporting systems at high volumes is solved by multiple innovations in sensor and backend packaging, deployment, integration and data searching, based on SISSDEN's consortium's extensive experience with "big data" approaches, high volume transactional and non-relational data systems.
- Share high-quality actionable information on a large scale. SISSDEN produces large amounts of intelligence on current threats and all of it is being shared with stakeholders and the larger community, at no cost to them, for the purposes of remediation or for early warning. The project distributes high-quality data feeds to the majority of the National CERTs in Europe, as well as worldwide, along with Law Enforcement Agencies, Internet providers, network owners and other vetted organisations fighting to defend their networks, SME customers, EU citizens and Internet users against continuous attacks.
- Provide objective situational awareness through metrics. Access to huge amounts of high-quality data on cyber threats: primarily obtained by the sensor network, but also contributed by the members of the SISSDEN consortium, provides metrics that offer objective, non-vendor biased overview of the threat landscape in the EU and individual member states.
- Create and publish a large scale curated reference data set. A significant subset of the data produced by SISSDEN is being made available to vetted researchers and Academia, addressing the clear and urgent need for large scale, high quality, and recent security datasets in order to improve or test defensive solutions.

6.2.2 Technical Architecture

Figure 6.2 below provides a simplified view of the SISSDEN technical architecture.

Components located at the EU datacentre include the Frontend Servers, Backend Servers and Utility Server pictured on the diagram. The sensor network consists of remote VPS Provider end points located at various VPS hosting providers (i.e. outside the EU datacentre), configured as transparent network tunnel endpoints forwarding traffic to the EU datacentre. SISSDEN collects attack data, such as network scans, spam email, malware binaries, brute force attacks, interactive attacker logins, etc.

6.2.2.1 Remote endpoint sensors (VPS)

Each remote endpoint sensor contains only the minimum amount of configuration and management capabilities required to securely participate as one end of a transparent network tunnel. They are configured to act as a long virtual Ethernet cable between the VPS and SISSDEN's local data centre frontend. At the Frontend in the EU Datacenter, a tunnel server terminates each transparent layer 2 Ethernet tunnel and delivers the Ethernet frames to an isolated, dedicated Virtual Local Area Network (VLAN).

6.2.2.2 Frontend servers

Traffic from the remote sensor endpoints are received by multiple types of honeypot systems, implemented as VMs, running on the EU Datacentre Frontend. Each honeypot VM emulates one or more potential vulnerabilities and collect data about attacks observed against those vulnerabilities. The honeypots have a standard configuration and standard data collection formats

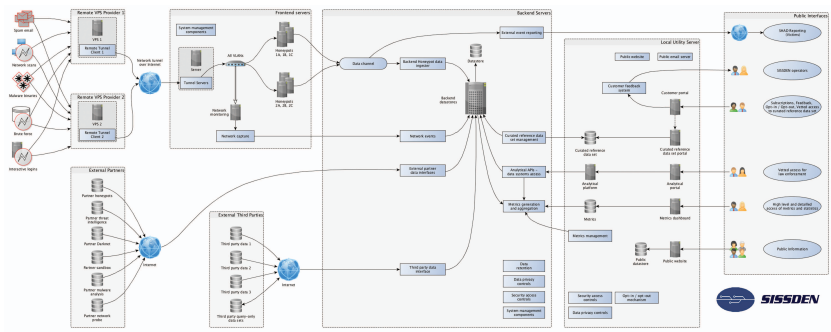

Figure 6.2 High-level architecture of the SISSDEN network.

enabled. Their data collection capabilities are complemented by network packet capture components (using solutions such as MMT and Snort) running on separate VM instances that listen to all traffic coming to them. SISSDEN system management components centrally manage all VM configuration, orchestration and operations.

Honeypot data and data from the network capture components are being ingested into the Backend datastores located at the Backend Servers at the EU Datacentre.

Tools like MMT and Snort are used to capture and analyse the network traffic. Snort allows identifying attacks using known attack signatures. MMT (Montimage's monitoring framework), adapted for SISSDEN, allows characterising malicious behaviour that corresponds to both known and unknown attacks. This information, referred to as CTI, is used by this monitoring framework for automating the real-time prevention and mitigation of attacks to an organisation (large or small) before they reach their network.

6.2.2.3 External partner and third-party systems

The data collected by the SISSDEN sensor network is supplemented by data from external systems operated by SISSDEN partners. These include separate honeypot networks, darknets, sandbox and malware analysis systems, threat intelligence platforms, etc. As with the sensor network, data from these systems is being ingested in various forms and stored in the Backend data stores.

To avoid unnecessary software development, SISSDEN makes use of and extends background partner systems, which aggregate data from multiple sources and provide a well-defined RESTful API for accessing normalized datasets.

6.2.2.4 Backend servers

Data from SISSDEN's various data collection systems is presented in multiple formats, such as live-streamed events, log files, PCAP files, and other file format data. Most of these data types are stored in their raw format in local data storage systems, at least for predetermined periods/repository size quotas, and some of the data types require parsing, normalization and ingesting into backend data indexes in support of free daily remediation report generation, high-value CTI, data analytics and ad-hoc querying.

6.2.2.5 External reporting system

One of the main purposes of the SISSDEN project is to collect Internet scale, timely security event data and make it available at no cost to vetted National CERTs, Network Owners and organizations who sign up for SISSDEN's free daily alerts.

The various sources of data collected by SISSDEN, such as honeypot and darknet data, malware analysis data, and botnet tracking information – as well as ingested external third party data sources – is being collected and stored locally in the SISSDEN backend. Each day, recipients who have voluntarily signed up for free reporting will receive by email multiple reports, covering different types of potentially malicious activity detected by SISSDEN on their nominated, verified IP/ASN/CIDR addresses.

On the other hand, SISSDEN BV will further provide real-time CTI, through a subscription service, to allow any organisation to block identifying cyber-attack campaigns before they reach their networks.

6.2.2.6 Utility server

Various analytics are being performed on the data collected by SISSDEN. An analytics platform is being extended, and hosted on the Utility Server. These analytics solutions provide additional insight into threats propagating in the Internet, pooling together partner resources dedicated to the project. In addition, metrics are being applied to the collected datasets to provide improved situational awareness. They can be used as a basis on which informed decisions can be made to mitigate threats. Curated reference datasets are also being made available to vetted researchers through the Utility Server. Interactions with the above are described in more details in this document and take place through the external interfaces illustrated in the diagram (with the exception of the analytics platform, which is only available to SISSDEN partners).

SISSDEN presents a number of systems to interact with the public and external partners. These include a Public website (mostly containing information about the project), email communication (reports), a Customer Portal, Metrics Dashboard, etc. Hosted on the Utility Server, these public facing systems include mechanisms to communicate with the consortium, sign up to request free of charge reports, gain access to the curated reference data set, provide customer feedback, and manage opt in/out and data privacy issues.

6.2.3 Concrete Examples

Two use cases, among many, have been selected to illustrate the real added-value of the CTI information that is provided.

6.2.3.1 Use Case 1: Targeted Cowrie attack that can be anticipated by the analysis of the traffic before it occurs

Targeted attacks are one of the emerging trends in cyber-security. Unlike conventional network scans and massive operations like spam and phishing, these attacks are generally answering the following criteria:

- They are focused on the assets of a single victim (private institution, government, critical infrastructure...) with objectives such as Data Exfiltration and Service Disruption.
- In the case of Data Exfiltration, it needs to be prepared and carried out after studying the infrastructure of the victim. The attackers will most probably put a lot of effort to hide their activity.
- In the case of Service Disruption, the attack is generally based on DDoS activity to disrupt the services and assets of the victim. This objective is normally achieved in a very short time (few minutes) and could be carried out repeatedly, thus generating an annoying service disruption, and consequently impact the victim's reputation. If the victim is a cyber-security company, the attack may take offline important security infrastructure (such as IDPS, honeypots and firewalls) and thus open the door to other attacks toward the protected zones (clients' assets, infrastructures...).

From a network traffic point of view, a targeted attack on honeypots looks like the curve shown in Figure 6.3. The spike shows when the targeted honeypot and/or its back-end system are hit. The graph shows the number of events registered in by the honeypot system which led to a 2-hour downtime of the honeypot system.

One can see the "normal traffic noise" before the attack and after the system has recovered.

Service suppliers (e.g., hospitals, media, power plants, control systems) cannot afford a 2-hour downtime. This class of attacks are able to disrupt the majority of infrastructures on the market. This has led the SISSDEN BV team to develop a DDoS resilient honeypot that will detect but not suffer from these attacks and therefore offer customers an improved security and uninterrupted threat analysis/monitoring.

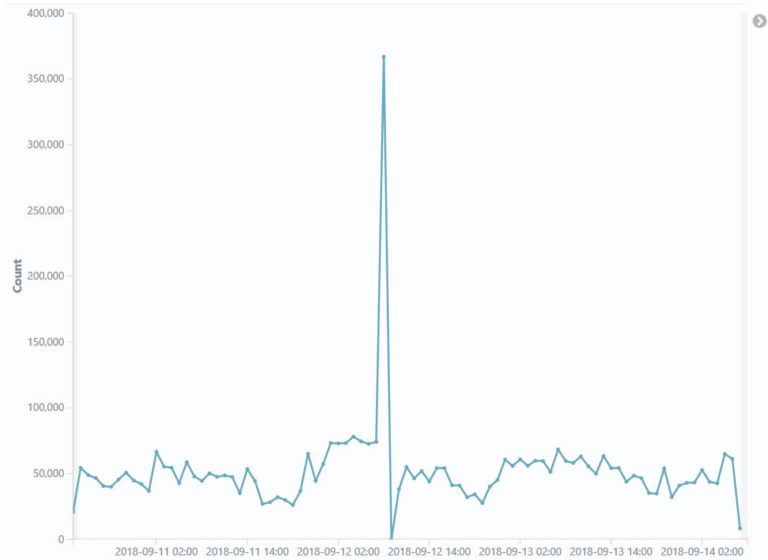

Figure 6.3 Shows the genesis of the attack over time (measures made each 5 minutes).

Furthermore, with the information that can be provided, the customers can prevent that their services do not go down in their own networks. For this, they can redirect or drop all the ingress traffic coming from the sources of this spike (using the IP addresses) and, if the attack starts occurring, set up another path for the egress traffic.

6.2.3.2 Use Case 2: Understanding the numbers – metrics

SISSDEN delivers realistic up-to-date metrics data and dashboards from its own sources that are compared and complemented with collated sources. SISSDEN categories are based on digital epidemiology and evidence-based practices as modelled from prior knowledge and research gained from other H2020 EU Projects: SAINT[2] and CyberROAD[3]. SISSDEN provided data can be used to make more informed decisions and improve security outcomes for clients. For instance, CTI data from SISSDEN and related sources found that in the first quarter of 2018 alone, the average enterprise faced:

[2]https://cordis.europa.eu/project/rcn/210229_en.html and https://project-saint.eu/
[3]https://cordis.europa.eu/project/rcn/188603_en.html and http://www.cyberroad-project.eu/

- 21.8% of all Website traffic that is due to bad bots (a 9.5% increase over the first quarter of 2017). For example, click fraud is a major threat especially for ISP's and enterprises, 1 out of 4 clicks are now fraudulent.
- 7,739 malware attacks (a 151% increase over the first quarter of 2017).
- 9,500 Botnet C&Cs (Command and Control servers) on 1,122 different networks (a 25% increase over the first quarter of 2017).
- 173 ransomware attacks (a 226% increase over the first quarter of 2017).
- 335 encrypted cyber-attacks (a 430% increase over the first quarter of 2017).
- 963 phishing attacks (a 15% year-over-year increase).
- 554 zero-day attacks (a 14% increase over 2017).
- 5,418,909,703 (5.4 billion) Web-based user accounts that have been compromised by 310 known or reported data breaches (a 40% increase over the first quarter of 2017).
- 40% of business and government networks in US and Europe shown evidence of DNS tunnelling.
- 75% of application DDoS, like HTTP-flooding, was in fact automated threats to Web applications mistakenly reported as DDoS.
- 73% of cyber-attacks focused on the cloud were directed at Web applications.
- 755 of 62,167 of the ASNs (autonomous systems) in routing system (1%) account for hosting, routing and trafficking 85% of all malicious activity.
- 13,935 total incidents are either route hijacks or outages. Over 10% of all ASNs were affected. 3,106 ASNs were a victim of at least one routing incident. 1,546 networks caused at least one incident in 2017 and already up by 20% in 2018.
- 90% of enterprises feel vulnerable to insider attacks, of which 47% are insiders wilfully causing harm and 51% are from insiders by accident; compromised credentials, negligence etc.

Ultimately analysing this type of metrics data by attack type, origin and region helps enterprises understand how cyber-attack trends are evolving. SISSDEN BV innovative AI approaches help in the timely prevention of these threats, remove false positives, help improve budget/resources prioritisation, and improve awareness with open source feeds.

6.3 Conclusion

Many security-oriented tools and services exist that provide or use CTI for the prevention, detection and response to threats. CTI is integrated natively into security products (i.e., appliances and software tools) or provided as a service for organisations' response teams. Among those that offer state-of-the-art Threat Intelligence solutions and services we have, for instance [2]: Anomali, ThreatConnect, ThreatQuotient, LookingGlass and EclecticIQ.

With respect to these offers, the SISSDEN project provides free feeds derived from its wide network of honeypots and darknets; and, the start-up, SISSDEN BV, provides original real-time actionable feeds complemented with information from other sources, that are not provided by these companies since they mainly rely on the existing open data that is analysed offline.

The innovation with respect to state-of-the-art market solutions provided by SISSDEN concerns the following:

- Ease of use and comprehensive threat indicators: SISSDEN relies on open standards (e.g., STIX/TAXII) and provides malicious-only IP addresses, subnets, URLs, threat ontology and ASNs.
- Trust in provided intelligence and accuracy: SISSDEN intelligence comes from malicious honeypot and darknet activity that contains no false positives.

The SISSDEN BV start-up further provides:

- Timely and Real Time: SISSDEN BV delivers CTI in real time (less than 1 minute) for effective blocking of attacks before they occur.
- CTI is correlated with information from other sources and using Deep Data and Artificial Intelligence-based analysis, increasing its value and extent.
- Removing complexity: SISSDEN BV allows for efficient use of security resources and provides shared threat intelligence and automated response.
- Modular and scalable: SISSDEN BV can serve different categories of customers: SMEs without security expertise or solutions, medium and large enterprises with their own solutions and security teams...

Acknowledgements

This work is performed within the SISSDEN Project with the support from the H2020 Programme of the European Commission, under Grant Agreement No 700176. It has been carried out by the partners involved in the project:

- Naukowa I Akademicka Siec Komputerowa, Poland
- Montimage EURL, France
- CyberDefcon LTD, United Kingdom and The Netherlands
- Universitaet des Saarlnades, Germany
- Deutsche Telekom AG, Germany
- Eclexys SAGL, Switzerland
- Poste Italiane – Societa per Azioni, Italy
- Stichting The Shadowserver Foundation Europe, The Netherlands.

References

[1] Bachar Wehbi, Edgardo Montes de Oca, Michel Bourdellès: Events-Based Security Monitoring Using MMT Tool. ICST 2012: 860–863

[2] Craug Lawson, Khushbu Pratap; "Market Guide for Security Threat Intelligence Products and Services" published 20 July 2017 by Gartner.