Blog Article #1
by ITI-CERTH

How SECANT stores and shares Cyber Threat Intelligence

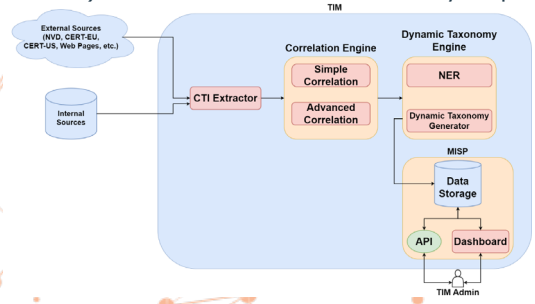
The SECANT project aims to deliver a holistic framework for cyber security risk assessment for enhancing the digital security, privacy, and personal data protection in complex ICT infrastructures by placing an automated threat detection form addressed to CERTs/CSIRTs that is capable of identifying threats and attacks, while promoting the situational security awareness as a priority within complex ICT infrastructures, such as the healthcare ecosystem.

Within the SECANT platform, particular emphasis is given to the collection, extraction and sharing of Cyber Threat Intelligence (CTI). To this end, the Threat Intelligence Module (TIM) is being developed whose main role is to identify, gather, enrich and share CTI data. This is implemented by performing a both manual and automatic collection of threats and vulnerabilities, by extracting and enriching this information by applying correlation techniques and by assigning cybersecurity taxonomies in a dynamic manner.

The TIM collects data from several internal and external sources. With regard to the internal sources of an organisation, the CTI data can be obtained from a variety of internal assets and events such as observable events that have happened on an organisation's internal network and hosts. In practice, these events can be anomalies in the internal organisations networks, web logs or system logs. External sources include sources such as vulnerability databases, CERT feeds, databases with Proof of Concept exploits, social media, forums, and relevant web pages from the Surface Web and the Dark Web.

The output of TIM consists of CTI and enriched CTI, mapped to taxonomies and will assist the parallel pillars of SECANT such as the vulnerability and risk assesment.

TIM comprises of four distinct sub-components which facilitate the difference functionalities, namely the CTI Extractor, the Correlation Engine, the Dynamic Taxonomy Engine, and the Malware Information Sharing Platform (see Figure). Each sub-module is responsible for specific functionalities including, data gathering, analysis, correlation, dynamic, taxonomy allocation and storage. TIM will communicate with SECANT's Interoperability Layer (IPL) which will provide data such as logs and alerts of existing devices within the organisation. IPL module will facilitate the collection of data from internal sources by gathering data directly from the devices (logs from legacy devices) and the Technical Vulnerability Impact Assessment module (TVIA). Apart from data from IPL, internal sources include the utilisation of honeypot instances for the collection of malicious data. In particular, the T-Pot56 solution will be used which is an all-in-one multi honeypot platform which comprises more than 20 different honeypots including honeypots concerning the healthcare domain (e.g., dicompot, medpot). In addition, the TIM module supports web crawling for gathering information from external sources such as relevant websites. Three main crawling functionalities are supported by the web crawler: the crawler follows links found in a web page and parses their content (general crawling), the crawler follows link found in web pages but parses only pages with content related to CTI (focused crawling) and the crawler imitates human user behaviour by employing a real web browser to navigate through the web (evasive behaviour). This functionality allows the crawler to scrape content that is dynamically loaded or different content that is served from websites when browsed by humans instead of crawling bots.ww



TIM high-level architecture

Furthermore, TIM leverages the MISP platform for gathering CTI from other external sources such CTI repositories (e.g., MalwareBazaar) as well as store the extracted CTI from both internal and external sources.
The gathered information is analysed in order to identify and extract IoCs and IoAs from the content. Subsequently, the extracted IoCs and IoAs compose the CTI entry that is enriched via simple and advanced correlation and stored on MISP as a MISP event. In order to avoid storing personal data, TIM filters the collected data leveraging a rule-based approach to either anonymise or pseudonymise any Personal Identifiable Information (PII). In addition to storing the output as MISP event, TIM supports the export in different formats such as PDF and STIX 2.1., thus facilitating the interoperability of the module. Finally, TIM supports the creation and dynamic adaptation of taxonomies for cyber-attacks using among others, ML-based techniques whenever new threat intelligence information is inserted to the platform.

In conclusion, cyber security intelligence data plays a critical role in safeguarding our digital world from cyber threats. By collecting, analysing, and interpreting vast amounts of data from various sources, cyber security professionals can identify potential threats, vulnerabilities, and attack patterns and take proactive measures to mitigate risks. However, cyber security intelligence data is only useful when it is relevant, accurate, and timely. Indeed, the SECANT consortium takes into account this reality and has a rigid and thorough strategy to gather, enrich and share accurate CTI.

This project has received funding from the European's Union's Horizon 2020 research and innovation program under grant agreement No. 101019645