The n6 platform

n6 was designed and developed entirely at CERT Polska as a platform for acquiring, processing and exchanging information regarding Internet threats. Currently, millions of security events are processed daily in an automated manner. The goal is efficient, reliable and fast delivery of large volumes of network incident data to interested parties: network owners, administrators and Internet Service Providers. The project disseminates information gathered from various security systems operated by security organizations, software vendors, independent researchers, etc.

The platform exchanges information on the source attacks in the form of URL addresses, domains, IP addresses or names of malicious software, and also information on special data (ZeuS configuration, etc.).

Malicious software, and also information on special data (ZeuS and fast delivery of large volumes of network incident data to accomplish that goal, the n6 API has been extended with additional attributes required to represent information present in all datasets used by the consortium.

To make it as easy as possible to make other datasets accessible using the same API without requiring direct integration with the n6 platform operated by NASK, the n6 SDK was developed and released as open source. A detailed tutorial on enabling n6 API in a dataset is provided in deliverable D3.2 of NECOMA.

n6 REST API & n6 SDK

One of the main design goals of the n6 API is to make integration as easy as possible from the client point of view. This property, combined with extensibility provided by JSON, ability to handle large amounts of data and filtering capability, makes the API a good candidate for exchange of homogeneous datasets and it was chosen as the basis of the common data exchange mechanism in the NECOMA project. In order to accomplish that goal, the n6 API has been extended with additional attributes required to represent information present in all datasets used by the consortium.

In the learning phase, for each sample a network behavior profile is created in the form of a sequence of flows identified by protocol, destination address and destination port and associated with a sequence of packet payload lengths. Then, using C&C/blacklist and legitimate servers whitelist, each flow is classified as either benign or C&C based on destination IP address.

In order to derive network behavior profiles, we cluster the initially classified samples with similar network behavior using agglomerative hierarchical clustering with Dynamic Time Warping (DTW) as the algorithm to compute the distance function, enabling matching of sequences with different lengths. For each cluster a network profile is built, which is essentially a sequence of flows similar to mentioned above but extended with weights for each packet payload length within a profile's associated with a sequence of packet payload lengths. Then, using C&C/blacklist and legitimate servers whitelist, each flow is classified as either benign or C&C based on destination IP address.

In the discovery phase, the distance between a sample containing flows directed to unknown servers (potentially C&C) and each cluster represented by its network behavior profile is computed using again DTW algorithm. This time, however, weights for each packet payload length within a profile's flows are taken into account. Finally, each unknown endpoint in the sample being tested is classified using flows from profile (clusters) closest to sample, matched with DTW algorithm.

The algorithm was tested on data from the n6 database. The following table shows the results for 248 IP addresses classified using knowledge gathered from a dataset containing 433 known C&C IP addresses and 740 benign addresses (resulting in 414 clusters, 50 best ones were used).

<table>
<thead>
<tr>
<th>Classification</th>
<th>Positive</th>
<th>False positive</th>
<th>True negative</th>
<th>C&amp;P</th>
<th>C&amp;I</th>
<th>C&amp;C</th>
<th>benign</th>
</tr>
</thead>
<tbody>
<tr>
<td>C&amp;C</td>
<td>56.08</td>
<td>25.00</td>
<td>57.13</td>
<td>0.63</td>
<td>81.47</td>
<td>58.00</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Graph-based cross-layer analysis

The graph-based analysis is a simple tool enabling exploration of the entire collection of datasets based on seed tokens: interesting values either identified by other analyzers, or obtained in a different way by the user of the system. The goal of the approach is to find as much information related to a given token as possible, while limiting the cost of such operation. Related information within different datasets can be modeled as a graph, with information from a single dataset represented as a node and links between nodes representing related information.

The algorithm proposed in deliverable D2.1 builds a graph of related information starting with a token. Our solution creates a graph of tokens and assigns a cost to each query, building the graph by selecting cheapest queries until a preset budget is consumed.

The proposed analysis has two separate usage cases. In interactive mode it is a tool used to explore the datasets, starting with a manually chosen token. In automatic mode the analyzer processes tokens provided by other analyzers or taken from some interesting datasets.

Automated rating and classification

NASK has developed several metrics of dataset usefulness, both manual (configurable or reputation-based) and automatically identifiable. The ratings are heavily used in the graph-based analysis to select the most useful data sources in the process of graph creation.

Another type of rating mechanism is proposed for the process of collecting external knowledge. In this case free APIs of research engines are used to assess the quality of search results provided for the tokens identified by the NECOMA platform. These results, coupled with the cost model of each engine, allows us to select the most cost-efficient choice of paid service capable of handling the large amount of queries generated for the tokens identified by the NECOMA platform with a good chance of collecting valuable information.

WORKPACKAGE 4: Case Studies

Malware campaign mitigation

NASK will demonstrate how the solutions developed in previous workpackages deal with a simulated new malware campaign, verifying each of the necessary steps: detection of individual incidents (attack attempts against servers, webpages turning malicious, etc.), correlation of the collected data showing a repeating pattern, identification of a global campaign, data enrichment, collection of external information and response to the threat.

The demonstration will be based on a snapshot of real threat datasets, showing all stages of analysis leading to the identification of malicious campaigns and enabling proper reaction. We are currently finalizing the development of the metrics measuring the effectiveness of our approach.

About NASK

In 1991, NASK connected Poland to the Internet. Since December 1993, NASK has been a research & development organisation and is a leading Polish data networks operator. We offer state-of-the-art telecommunications and data solutions to business, administration and academic customers in Poland as well as the Polish national registry of Internet names in the .pl domain.

Scientific Activity

As a research institute, NASK carries out numerous scientific and research & development activities. Projects centre on telecommunications & data quality (QoS – Quality of Service), security of IT systems and automatic identification methods. NASK is an active member of many international organizations and associations (FIRST, CENTR, TERENA, RIPE) and participates in national and European Union projects.

NASK’s participation in the NECOMA project is a joint activity of two projects NASKIS (joined during the project by the Software Development Department).

CERT Polska

A part of the NASK organisation, CERT Polska is a Computer Emergency Response Team operated by NASK that handles incidents related to the .pl namespace. A part of the team’s work is focused on analyzing new detection and abnormal methods and developing tools to aid this process.

Network and Information Security Methods Team

A part of the NASK Research Division dealing with security problems, the NSIM team focuses on researching new detection and analysis methods and developing tools to aid this process.

The team’s user theoretical and exploratory approach complements the operational experience and focus of CERT Polska.