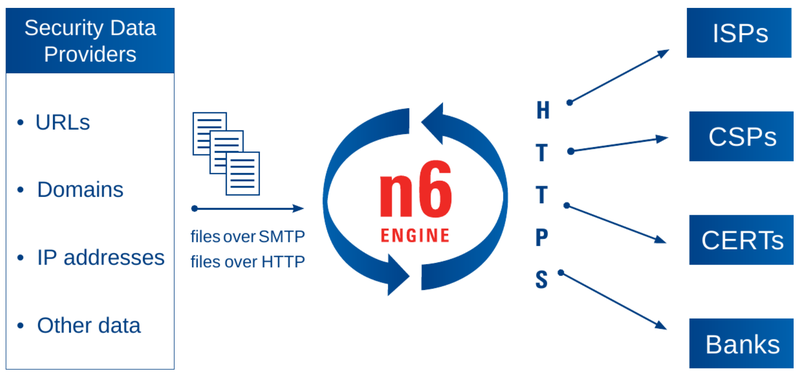
WORKPACKAGE 1: Threat Data

The n6 platform

n6 was designed and developed entirely at CERT Polska as a platform for acquisition, processing and exchange of information reagrding 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 config etc.).



The n6 platform

Data sources

The n6 platform handles many different types of data feeds, including malicious URL addresses, infected hosts (bots), C&C servers, phishing, spam, scanning, DDoS, brute force attacks, open-proxies and open-resolvers. Methods used currently to obtain the data or added as part of this workpackage include:

- direct botnet controller observation
- botnet dropzone observation
- monitoring traffic on a proxy-server
- active crawl of a peer-to-peer botnet
- passive listening to traffic in a peer-to-peer botnet
- data obtained from sinkhole
- results from behavioral analysis
- interaction with honeypots, both client and server side
- monitoring of traffic collected by darknet
- reports from anti-virus systems
- reports from intrusion detection and prevention systems
- reports from web application firewalls

For NECOMA NASK is integrating several new data sources running in-house into the platform, including:

A **sandbox** data source, based on the Cuckoo sandbox solution. The sandbox will provide information about connections initiated by the monitored malware, as well as information about the malware itself, using the malware repository operated by NASK.

A **client honeypot** data source. Possible solutions include the *Thug* client honeypot system and/or a more complex solution called the HoneySpider Network, developed by and operated by NASK (in international cooperation).

An **early warning system** data source and a **darknet** data source. Early warning alerts and darknet data will be provided respectively by the Arakis 2.0 early warning system developed at NASK and its darknet sensor.

Automated knowledge collection

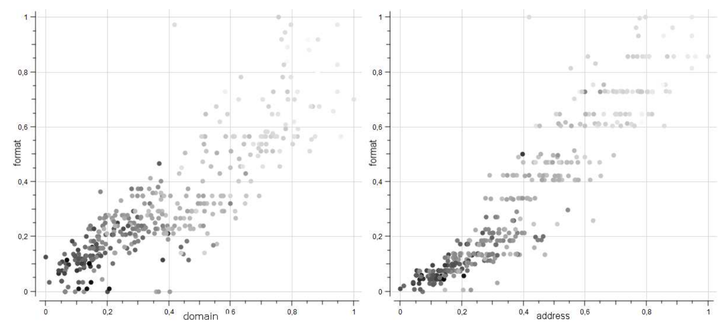
The automated knowledge collection system concieved at NASK will provide security-related information collected from the Internet. The mechanism will use tokens extracted from available threat data to search for associated knowledge using existing search engines.

The study of several available search engines regarding their usability in this scenario has been performed and possible ways of extracting simple but useful tokens have been proposed. Currently the collection mechanism is being implemented and research on utility of some less obvious token types is performed.

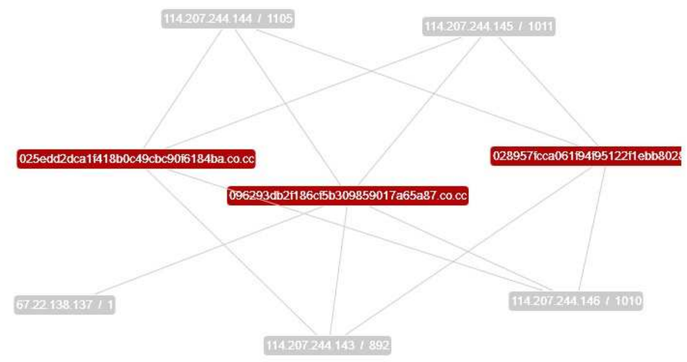
WORKPACKAGE 2: Threat analysis

Data analysis - current research topics

Threat data classification. Applying a Support Vector Machine to a classification problem consists of two steps: training and prediction. During training process, the Support Vector Machine takes as input a dataset in which each example is a fixed-length vector.



Support Vector Machine classification of malware using polynomial kernel core function

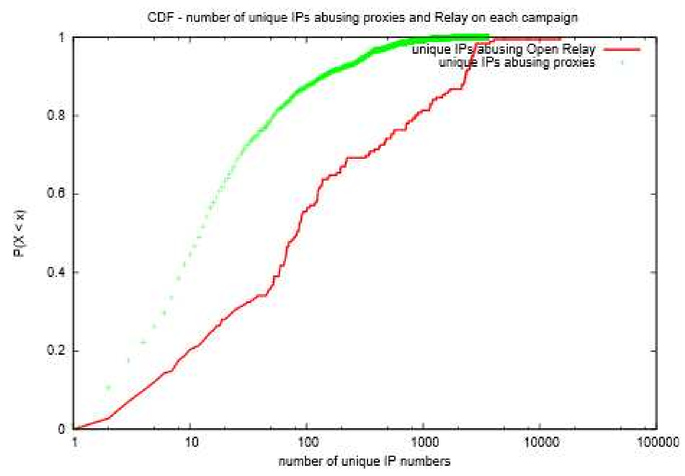


Building black list based on events in the n6 database

Threat Campaigns Detection We are going to elaborate time correlations between spam campaigns and malware propagation occur into the network. At first grouping of spam into different classes is required to perform efficient analysis. The classes include:

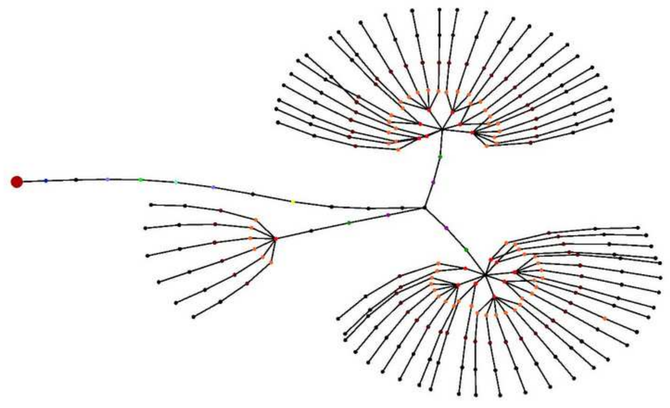
- Content SPAM
- Body spamming
- Anchor text spamming
- Link spamming
- Cloaking and redirection
- Title spamming
- Meta-tags spamming
- URL spamming
- Click SPAM

Important features extracted from the n6 and spam messages include time and date, type, IP address and domain name.



Unique IPs abusing proxies and relay on each campaign

Sequential Pattern Tree After spam campaigns are identified, we apply association rule mining algorithms to determine co-occurrence of campaign attributes that unveil different spamming strategies. In particular, we are going to find time correlations between spam campaigns and malicious events.



Sequential Pattern Tree

The presented methodology for characterizing spamming strategies is based on the grouping of spams into spam campaigns and then detecting invariant and co-occurrence patterns among them. Our technique builds a Frequent Pattern Tree using relevant features extracted from n6 database and spam messages. Based on that, messages that share a common frequent path in the tree and differ only on infrequent features are grouped into campaigns.

Automated rating and classification

This separate research activity at NASK deals with the varying quality of knowledge gathered from external sources by the automatic system developed in WP1. The goal of the research is to find methods to automatically identify low-quality sources, rate the relevance of information to the observed data and group the resources into different classes (e.g. CVEs, popular press mentions, statistics, etc).

WORKPACKAGE 3: Cyberdefense

Involvement of NASK in this workpackage is rather limited. Basing on experience from previous projects we will take a small part in the development of information exchange mechanisms. NASK will also provide through this mechanism any useful threat signatures generated by data analysis or provided by other NASK systems.

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.

About NASK

In 1991, NASK connected Poland to the Internet. Since December 1993, NASK has been a research & development organization and a leading Polish data networks operator. We offer state-of-the-art telecommunications and data solutions to business, administration and academic customers. NASK is also 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 biometric 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 groups within NASK (joined during the project by the Software Development Department):

CERT Polska A part of the NASK organization, 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 researching new detection and analysis 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 NISM team cooperates often with CERT Polska in security-related research projects. The team's more theoretical and exploratory approach complements the CERT Polska's operational experience and focus.