SEVENTH FRAMEWORK PROGRAMME

Information & Communication Technologies
ICT

Cooperation Programme

Nippon-European Cyberdefense-Oriented Multilayer threat Analysis[†]

Deliverable D5.6A
User and Contributor Guide for NECOMA Results

Contractual Date of Delivery	July 17, 2016
Actual Date of Delivery	July 17, 2016
Deliverable Dissemination Level	Public
Editors	Jouni Viinikka and Gregory Blanc
Contributors	All *NECOMA* partners

The *NECOMA* consortium consists of:

Institut Mines-Telecom	Coordinator	France
ATOS SPAIN SA	Principal Contractor	Spain
FORTH-ICS	Principal Contractor	Greece
NASK	Principal Contractor	Poland
6CURE SAS	Principal Contractor	France
Nara Institute of Science and Technology	Coordinator	Japan
IIJ - Innovation Institute	Principal Contractor	Japan
National Institute of Informatics	Principal Contractor	Japan
Keio University	Principal Contractor	Japan
The University of Tokyo	Principal Contractor	Japan

[†] The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7-ICT-2013-EU-Japan) under grant agreement n° 608533 and the Strategic International Collaborative R&D Promotion Project of the Ministry of Internal Affairs and Communication, Japan.

Contents

1 Introduction

The *NECOMA* project exploitation plan [7] being a confidential document focusing on the partner's use of project results, this document is a public appendix, oriented to guide third parties in the use of and contribution to the *NECOMA* results.

The document provides two different, albeit overlapping, views to *NECOMA* results. First, we provide insights to different user categories such as researchers or enterprises in Chap. 2. In Chap. 3, we orient different types of contributors – threat data providers and users – towards relevant *NECOMA* results. In addition, Chap. 4 orients the reader towards different lessons learned during the project.

We refer only to publicly available deliverables and provide direct links to documents, websites, and source code repositories to ease the access to information and results.

2 User Guide

This chapter looks at the *NECOMA* project results from the point of view of an entity looking to make use of *NECOMA* project results. We provide an overview of potentially interesting results for different types of users, researchers, students, Internet Service Providers (ISP), SMEs, industry, and CERTs in the following sections. The content overlaps for some categories, as we believe different categories may be interested in the same results.

Table 2.1 lists the results identified in the exploitation plan and provides pointers to deliverables with more details, including URLs towards online resources, such as websites or github repositories. The columns Res, Stu, SME, Ind, ISP, CERT are used to indicate whether we think that Researches, Students, SMEs, large enterprises/industrials, Internet Service Providers, and CERT-like organizations might be interested in a given result, respectively.

Result	Owner	Deliv.	Contact	Res	Stu	SME	Ind	ISP	CERT
High Performance Phishing Detection	ATOS	D2.1, §2.3.4; D2.2	Dawid Machniki	x			x	x	x
DNS-based Detection for Cloud	ATOS	D2.2, §6.1	Dawid Machniki	x				x	
SDN test platform	IMT	D4.1, §6.1	Gregory Blanc	x	x	x	x	x	x
Accurate honeypot results	FORTH	D2.1, §3.2	FORTH	x		x	x	x	x
n6 SDK	NASK	D1.4, §2.1	NASK	x		x			x
n6 stream API	NASK		NASK			x	x	x	x
Analysis modules	NASK	D2.1, §2.4.3, 2.4.4; D2.2, §3.3.1, 3.4.2	NASK						x
Dataset rating methods	NASK	D2.1, §3.3; D2.2, §4	Pawe Pawliski				x	x	x
Router reconfig. tool	6cure	D3.5, §2.1;	Jouni Viinikka	x			x	x	x
Router reconfig. tool integration	6cure		Jouni Viinikka					x	
Detection of synchr. sources	6cure	D2.1, §2.1.2; D2.2, §3.1.6	Jouni Viinikka	x			x	x	x
SDN-based DDoS Mitigation	NAIST	D3.5, §3.3	Kazuya Okada	x				x	
LISP-based DDoS Mitigation	NAIST	D3.5, §3.4	Kazuya Okada	x				x	
Drive-by-download prevention	NAIST		Kazuya Okada	x			x		
Agurim Tools and Dataset	IIJ	D1.4, §3.1.1.2	Kenjiro Cho	x	x		x	x	
Tamias Distributed Storage	IIJ		Kenjiro Cho	x					
Anomaly Taxonomy	NII	D2.1, §3.1	Johan Mazel	x	x		x	x	
Programmable IX	Keio	D3.5, §3.3; D4.1, §2.3.2	Keio					x	
Sample Program for Smartphone	Keio		Keio		x	x			
Phishing Education	UT	D2.1, §2.3.2; D2.2, §6.3	Daisuke Miyamoto	x					
MATATABI Threat Detection	UT	D2.1, §4.2; D2.2, §2.2	Yuji Sekiya	x			x	x	x

Table 2.1: Summary of *NECOMA* results. Those without links to deliverables have not been presented in public deliverables. The columns on the right hand side indicate the potentially interested user categories as used in this document

Dataset	Owner	Deliv.	API	Contact	Type	Format	Avail
WIDE-TRANSIT traces w/ payload	IIJ/WIDE	D2.1, §3.1.1.1	N/A	Kenjiro Cho (IIJ/WIDE)	traffic	pcap	R, A
WIDE-TRANSIT Aggregated Flow Data	IIJ/WIDE	D2.1, §3.1.1.2		Kenjiro Cho	traffic	aguri2	R, A
Packet traces from a university	IIJ/WIDE	D2.1, §3.1.1.3		Kenjiro Cho	traffic	pcap	R
Netflow data from universities	UT	D2.1, §3.1.1.4	MATATABI, n6	Yuji Sekiya	traffic	nfdump	
sFlow data from Internet backbone	UT	D2.1, §3.1.1.5	MATATABI, n6	Yuji Sekiya	traffic	sflowtool text	
sFlow data from a public cloud	UT	D2.1, §3.1.1.6	MATATABI, n6	Yuji Sekiya	traffic	sflowtool text	
sFlow data from universities	UT	D2.1, §3.1.1.7	MATATABI, n6	Yuji Sekiya	traffic	sflowtool text	
DNS query data from a WIDE DNS Server	UT	D2.1, §3.1.2.1	MATATABI, n6	Yuji Sekiya	dns	pcap	
DNS query logs from cache resolver	UT	D2.1, §3.1.2.2	MATATABI, n6	Yuji Sekiya	dns	log	
DNS query data from cache resolver	UT	D2.1, §3.1.2.3	MATATABI, n6	Yuji Sekiya	dns	pcap	
DNS query logs from authoritative servers	UT	D2.1, §3.1.2.4	MATATABI, n6	Yuji Sekiya	dns	log	
DNS query data from authoritative servers	UT	D2.1, §3.1.2.5	MATATABI, n6	Yuji Sekiya	dns	pcap	
DNS query data from M-root DNS in DITL	KEIO	D2.1, §3.1.2.6	ssh	Akira Kato	dns	pcap	
OSPF topology from the WIDE backbone	IIJ/WIDE	D2.1, §3.1.3.1	HTTP GET	Kenjiro Cho	topo	text	R
iBGP datasets from WIDE (AS2500)	IIJ/WIDE	D2.1, §3.1.3.1	download	Kenjiro Cho	topo	MRT	R
Darknet traffic traces from NII	NII	D2.1, §3.1.4.1		Kensuke Fukuda	telescope	pcap	
Port scans detected by ARAKIS	NASK	D2.1, §3.1.5.1	n6	CERT Polska/NASK	early warn	JSON/n6	*
Attacks detected by ARAKIS	NASK	D2.1, §3.1.5.2	n6	CERT Polska/NASK	early warn	JSON/n6	*
NEMU Malware Data	FORTH	D2.1, §3.1.5.3	n6	Thanasis Petsas	early warn	mysql	
BotHunter Botnet Data	FORTH	D2.1, §3.1.5.4	n6	Thanasis Petsas	early warn	mysql	
DNS reflection attack data	NAIST	D2.1, §3.1.5.5		Kazuya Okada	early warn	hive	
NTP reflection attack data	NAIST	D2.1, §3.1.5.6		Kazuya Okada	early warn	sflow, netflow	
KEIO spam	KEIO	D2.1, §3.2.11		Akira Kato	mail	RFC821	
UT spam	UT	D2.1, §3.2.1.1		Yuji Sekiya	mail	RFC821	
Phishing URLs	UT	D2.1, §3.2.2.1		Daisuke Miyamoto	web	postgresql, hive	
Phishing content	UT	D2.1, §3.2.2.2		Daisuke Miyamoto	web	postgresql, hive, media	
SSL server response	IMT	D2.1, §3.2.2.3		Gregory Blanc	web	parsifal	
Credibility assessment of websites	UT	D2.1, §3.2.3.1		Daisuke Miyamoto	user behavior	postgresql, hive, media	
Data from sinkholes	NASK	D2.1, §3.2.4.1	n6	CERT Polska/NASK	sinkhole	JSON/n6	*
Network connection attempts by malware	NASK	D2.1, §3.2.5.1	n6	CERT Polska / NASK	client honey	JSON/n6	*
Peer-to-peer bot list	NASK	D2.1, §3.2.5.2	NASK	CERT Polska/NASK	client honey	JSON/n6	*
Malicious URLs from multiple sources	NASK	D2.1, §3.2.6.1	n6	CERT Polska/NASK	malic. URL	JSON/n6	*

Table 2.2: *NECOMA* datasets. The availability column indicates R for available upon request, A for anonymized data publicly available, and * for data NASK shares with AS owners for the entries concerning their own AS only (for remediation purposes)

2.1 Researchers

First of all, for researchers in the domain of threat data collection, threat analysis, and/or countermeasure application, the project website[1] hosts many publications made by the project team.

Those who are in the possession of datasets they wish to share, we refer you to the Sect. 3.1 for instructions and pointers towards the data exchange API defined by *NECOMA*. In brief, the API allows the data owners to maintain control over the data, while providing controlled access to it to third parties. Another advantage is that your data, of course depending on its actual contents, is directly usable by existing *NECOMA* analysis modules and platforms.

For those who wish to be able to use data collected by the project, the Table 2.2 lists the different datasets. The table describes, when relevent, the data format, the access API, and the availability of the data for third parties. Overall, we suggest considering implementing your analysis tools the n6 API for data access, allowing your tools an easier consumption of data collected by the *NECOMA* project and also by any other entity sharing their data using the *NECOMA* interfaces.

For those wishing to use or build on tools developed by the project, part of the tools have been released as open source. Among those, some continue to be maintained by the developers, while some are released "as is" for interested parties to take on their developement.

The Table 2.1 list the results, including tools, the project partners have identified as having highest exploitation potential. Furthermore, the public deliverables D2.1 Threat Analysis [1] and D3.5 Countermeasure Application - Results [5] provide a complete list of analysis tools and defense mechanisms, respectively, developed in the project.

If you wish to implement your own analysis platform, the information available in the deliverable D2.2 Threat Analysis Platform, Chap. 2 [3] describes the architecture and implementation of the *NECOMA* threat analysis platform.

If interested in threat data exchange, deliverables D1.4 Threat Data Final Report, Sect. 2.1 [6] and D3.3 Security Information Exchange - Results [4] discuss existing data exchange formats, document design choices made by the *NECOMA* project, and describe our data exchange formats.

Automated dataset rating and classification mechanisms may be of interest for researchers and are described in the deliverables D2.1 Threat Analysis, Chap. 3 [1] and D2.2 Threat Analysis Platform, Chap. 4 [3].

Different deliverables include discussion on lessons learned. That experience is likely valuable input as well for researchers in those domains.

[1] http://www.necoma-project.eu/publications/

Chapter 4 provides pointers to lessons learned in data collection, data exchange, and data analysis.

2.2 Students

For students, the deliverable D2.1 Threat Analysis [1] provides an overview of different threat analysis methods used in the project and it can serve as a starting point in familiarising with the domain. Many of the analysis modules described in the document can also serve as concrete examples of where different theoretical concepts, such as machine learning algorithms, are used in practice.

Similarly, the deliverable D3.1 Policy Enforcement Point Survey [2] provides an overview of existing policy enforcement points for students seeking introductory material for countermeasure enforcement. Both of these deliverables are amongst the most downloaded files from the project website.

2.3 SMEs

For small and medium enterpises, the interesting results are likely to be highly dependent on the domain of the SME. If you are an actor in the threat analysis domain, the deliverables D1.4 Threat Data Final Report [6], D2.1 Threat Analysis [1], D2.2 Threat Analysis Platform [3] are a good starting point. If an actor in the countermeasure application domain, the deliveable D3.5 Countermeasure Application - Results is a good starting point.

2.4 Industry/Large Enterprises

Industrials and large enterprises might be interested in taking up some *NECOMA* results, especially tools and mechanisms, for further development. The deliverable D2.1 Threat Analysis [1] provides an overview of different threat analysis modules. The deliverable D2.2 Threat Analysis Platform [3] provides additional information on those modules, especially in terms of experience gained in implementing them. Finally, the deliverable D3.5 Countermeasure Application - Results provides a description of the developed defense mechanisms. For the results the *NECOMA* partners have considered most interesting, the owners and contacts are listed in Table 2.1.

2.5 ISPs

For an Internet Service Provider, the analysis methods and corresponding tools developed in the project can provide interesting insights into their own

traffic. We suggest the deliverable D2.1 Threat Analysis [1] as a starting point.

An ISP might be interested in sharing data collected with organisations such as national CERTs or their peers and thus be interested in results related to data sharing and data access APIs, described in Sect. 3.1 of this document and detailed in deliverables D1.4 Threat Data Final Report, Sect. 2.1 [6] and D3.3 Security Information Exchange - Results [4].

If you wish to implement your own analysis platform, the information available in the deliverable D2.2 Threat Analysis Platform, Chap. 2 [3] describes the architecture and implementation of the *NECOMA* threat analysis platform.

Overall, making your data accessible through the n6 API and making your tools able to fetch/receive data via n6, eases the use of data produced by *NECOMA* partners during the project and after the project, and would allow feeding your data into the *NECOMA* threat analysis platform.

2.6 CERTs

A CERT or similar organization might act as a data provider, exactly as was the case of *NECOMA* partner NASK (and thus CERT Polska). If your intention were to provide data to *NECOMA* threat analysis platforms, we refer you to Sect. 3.1. The main interest in exposing your data using the proposed data exchange mechanisms is the usability of your data by *NECOMA* threat analysis platforms and tools.

If implementing your own analysis platform, the information available in the deliverable D2.2 Threat Analysis Platform, Chap. 2 [3] describes the architecture and implementation of *NECOMA* threat analysis platform.

If interested in threat data exchange, deliverables D1.4 Threat Data Final Report, Sect. 2.1 [6] and D3.3 Security Information Exchange - Results [4] discuss existing data exchange formats, document design choices made by the *NECOMA* project, and describe our data exchange formats.

Automated dataset rating and classification mechanisms may be of interest for CERTs and are described in the deliverables D2.1 Threat Analysis, Chap. 3 [1] and D2.2 Threat Analysis Platform, Chap. 4 [3].

Threat metrics are discussed in the deliverable D2.2 Threat Analysis Platform, Chap. 5 [3].

3 Contributor Guide

This chapter looks at the *NECOMA* project results from the point of view of an entity looking to contribute and participate, with a focus on threat data collection, and analysis tools and platforms. We provide two views, one for *data providers* who wish to make their data available to the analysis modules and/or platforms of *NECOMA*, and another for *data users*, who wish to be able to use the data offered by the mechanisms defined and developed in *NECOMA*.

3.1 Data Providers

We consider equipment vendors, service providers and dataset owners as potential data providers. Vendors and service providers in the sense that they can make their equipment (IDS, firewall, honeypot, traffic analysis system, etc.) or service compatible with the security information exchange format defined by the project. Once compatible, the service providers can share data directly and the vendors can enable their customers to share threat data to any data collection entity compatible with the *NECOMA* sharing interface.

Similarly, any data set owner can allow existing threat analysis platforms, analysis modules and/or defense mechanisms compatible with the *NECOMA* data sharing interface to access data they own and wish to share access to via the interfaces defined by the *NECOMA* project.

The n6 API is a key element in sharing datasets. The API and its implementation on datasets is described in the deliverables

- D1.4 Threat Data Final Report, Sect. 2.1.2 [6], and
- D3.3 Security Information Exchange - Results, Sect. 3.1 [4].

Information about concrete implementations in the *NECOMA* project of the n6 API for datasets is provided in the deliverables

- D1.4 Threat Data Final Report, Sect. 2.1.3 [6], and
- D3.3 Security Information Exchange - Results, Sect. 4.1 [4].

3.2 Data Users

We consider an entity wishing to use threat data collected in *NECOMA* data sets and/or analysis results produced by the *NECOMA* analysis tools and/or platforms for further processing as a data user. Further processing can mean, for example, analysis or use as input for a defense mechanism.

As for data providers, the key element is the n6 API. If your analysis module uses the n6 API to fetch analysis information, your module is able to use data shared by *NECOMA* partners and/or any other entity implementing the n6 interface to their data.

3.2.1 Data Consumer

For a data consumer, the API is described in the deliverable

- D1.4 Threat Data Final Report, Sect. 2.1.2 [6]

and feedback on its usage from a data consumer's point of view is provided in the deliverable

- D3.3 Security Information Exchange - Results, Sect. 3.3 [4].

3.2.2 Machine-to-Human Communication

If an analysis or defense mechanisms wishes to communicate its results towards humans, the *NECOMA* machine-to-human interface is NECOMAtter, described in the deliverable

- D3.3 Security Information Exchange - Results, Chap. 2 [4].

3.2.3 Adding Analysis Modules on MATATABI

MATATABI is an implementation of the *NECOMA* threat analysis platform, running on Apache Hadoop. If you wish to contribute to that platform, first of all, we suggest you to contact the MATATABI owners directly, see Table 2.1 for contact information.

The overall design of the *NECOMA* threat analysis platform is described in the deliverables

- D2.1 Threat Analysis, Sect. 4.1 [1] and
- D2.2 Threat Analysis Platform, Sect. 2.1 [3],

and the MATATABI platform itself is described in the same deliverables, in Sect. 4.2 and 2.2, respectively.

4 Lessons Learned

Overall, the data collection and sharing efforts made during the project have confirmed the impression we had already before the project: allowing public access to datasets with real world data is extremely complicated. There are many reasons for this, privacy and confidentiality concerns not being the least.

Furthermore, we believe that the datasets are valid and useful only for a very limited amount of time. In order for them to be truly useful, they need to be maintained actively, as the threat landscape, attacks, and the normal activity captured in the datasets evolve so quickly. For example, the Lincoln Lab experiments widely used by the intrusion detection community have been outdated and invalidated for a long time. Since they are publicly available, they are unfortunately still being used by many researchers, producing experimentations that do not tell anything about the real world detection capability of an intrusion detection system.

Several deliverables report more specific experience gathered and lessons learned during the project for different topics or discuss the existing approach and our design choices:

- The experience gathered while implementing the analysis modules is documented in the deliverable D2.2 Threat Analysis Platform, Chap. 3 [3].
- For security information exchange, the existing formats are discussed in D1.4 Threat Data Final Report, Sect. 2.1.1 [6]. The design choices for our machine-to-machine mechanism n6 have been presented in the same deliverable and in the same section. The design rationale for our machine-to-human mechanism NECOMAtter is presented in D3.3 Security Information Exchange - Results, Sect. 2.1 [4].
- The experience gathered on implementing the security information exchange mechanisms (n6 API) on datasets is described in D1.4 Threat

Data Final Report, Sect. 2.1.3 [6] and D3.3 Security Information Exchange - Results, Sect. 4.1 [4].

- The experience gathered on using the n6 API for querying information is described in D3.3 Security Information Exchange - Results, Sect. 3.3 [4].

Bibliography

[1] NECOMA Consortium. Deliverable D2.1: Threat Analysis. Technical report, Nov. 2014.

[2] NECOMA Consortium. Deliverable D3.1: Policy Enforcement Point Survey. Technical report, Nov. 2014.

[3] NECOMA Consortium. Deliverable D2.2: Threat Analysis Platform. Technical report, Nov. 2015.

[4] NECOMA Consortium. Deliverable D3.3: Security Information Exchange - Results. Technical report, May 2015.

[5] NECOMA Consortium. Deliverable D3.5: Countermeasure Application - Results. Technical report, Nov. 2015.

[6] NECOMA Consortium. Deliverable D1.4: Threat Data Final Report. Technical report, Apr. 2016.

[7] NECOMA Consortium. Deliverable D5.6: Exploitation Plan. Technical report, March 2016. (Confidential).