SEVENTH FRAMEWORK PROGRAMME

Information & Communication Technologies
ICT

Cooperation Programme

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

Deliverable D5.5: Final report on the use and dissemination of knowledge

Contractual Date of Delivery	March 31, 2016
Actual Date of Delivery	April 20, 2016
Deliverable Dissemination Level	Public
Editor	Daisuke Miyamoto
Contributors	All *NECOMA* partners

The *NECOMA* consortium consists of:

Institut Mines-Telecom	Coordinator	France
ATOS SPAIN SA	Principal Contractor	Spain
FORTH	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.

Contents

1 Introduction

The objective of Workpackage 5 is to ensure the spread of information about the *NECOMA* project by disseminating its output through different channels, such that they reach different types of audiences.

The audiences for the dissemination activities include the Ministry of Internal Affairs and Communications of Japan, and the European Commission. Furthermore, it is also intended for a wide range of targets, as the *NECOMA* consortium will promote the project's results in the scientific and industrial domain. The aim of the dissemination activities is to obtain feedback on the quality of the work carried out in the *NECOMA* project, as well as to raise the awareness of our contribution among practitioner communities.

We have thus planned, from the very beginning of the project, appropriate measures to make an effective and timely dissemination to academic and industrial domains, as well as any potential users working in this field.

In this respect, we have also published many research papers, presentations, proposals toward international standards, software, and datasets. The *NECOMA* project regards these activities as important for the dissemination of our knowledge, and essential for exploitation by many stakeholders.

The purpose of this Final Dissemination Report is to report all the dissemination activities that have been developed, implemented and applied during the *NECOMA* lifetime. These are divided into the following chapters: Chapter 2 provides the list of the published papers in the *NECOMA* project. Our available datasets and software are respectively introduced in Chapter 3 and 4. The dissemination through standards activities is explained in Chapter 5. *NECOMA* has also conducted diverse collaborative activities, including plenary meetings, teleconferences, and student/researcher exchanges as detailed in Chapter 6. Chapter 7 describes a set of short videos that demonstrate the case studies, and other types of dissemination activities are summarized in Chapter 8. We finally conclude the document in Chapter 9.

2 Research papers and other publications

Following is the list of peer-reviewed papers, posters and presentations that were presented or published in the period of June 2013–March 2016.

2.1 Research papers

1. Daisuke Miyamoto, Ryo Nakamura, Yuji Sekiya, Takeshi Takahashi. **Offloading Smartphone Firewalling Using OpenFlow-capable Wireless Access Points**. In *Proceedings of IEEE International Conference on Pervasive Computing and Communications (PerCom 2016)*. Sydney, Australia, March 2016. [pdf]

2. Daisuke Miyamoto, Yasuhiro Yamamoto, Masaya Nakayama. **Text mining-based Approach for Estimating Vulnerability Score**. In *Proceedings of the 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS 2015)*. Kyoto, Japan, November 2015. [pdf]

3. Ayumu Hirata, Daisuke Miyamoto, Masaya Nakayama, Hiroshi Esaki. **INTERCEPT+: SDN Support for Live Migration-based Honeypots**. In *Proceedings of the 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS 2015)*. Kyoto, Japan, November 2015. [pdf]

4. Panagiotis Papadopoulos, Thanasis Petsas, Giorgos Christou and Giorgos Vasiliadis. **MAD: A Middleware Framework for Multi-Step Attack Detection**. In *Proceedings of the 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS 2015)*. Kyoto, Japan, November 2015. [pdf]

5. Pernelle Mensah, Gregory Blanc, Kazuya Okada, Daisuke Miyamoto, Youki Kadobayashi. **AJNA: Anti-Phishing JS-based Visual Analysis,**

to Mitigate Users' Excessive Trust in SSL/TLS. In *Proceedings of the 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS 2015)*. Kyoto, Japan, November 2015. [pdf]

6. Daisuke Miyamoto, Gregory Blanc, Youki Kadobayashi. **Eye Can Tell: On the Correlation between Eye Movement and Phishing Identification**. In *Proceedings of the 22nd International Conference on Neural Information Processing (ICONIP)*. Istanbul, Turkey, November 2015. [pdf]

7. Evangelos Ladakis, Giorgos Vasiliadis, Michalis Polychronakis, Sotiris Ioannidis, and Georgios Portokalidis. **GPU-Disasm: A GPU-based x86 Disassembler**. In *Proceedings of the 18th Information Security Conference (ISC)*. Trondheim, Norway, September 2015. [pdf]

8. Iasonas Polakis, Michalis Diamantaris, Thanasis Petsas, Federico Maggi, and Sotiris Ioannidis. **Powerslave: Analyzing the Energy Consumption of Mobile Antivirus Software**. In *Proceedings of the 12th Conference on Detection of Intrusions and Malware and Vulnerability Assessment (DIMVA 2015)*. Milano, Italy, July 2015. [pdf]

9. Paweł Pawliński, Adam Kozakiewicz. **Lowering Cost of Data Exchange for Analysis and Defence**. In *Proceedings of the Coordinating Attack Response at Internet Scale (CARIS) Workshop*. Berlin, Germany, June 2015. [pdf]

10. Michał Kruczkowski, Ewa Niewiadomska-Szynkiewicz, Adam Kozakiewicz. **Cross-Layer Analysis of Malware Datasets for Malicious Campaign Identification**. In *Proceedings of the International Conference on Military Communications and Information Systems (ICMCIS 2015)*, Cracow, Poland, May 2015.

11. Romain Fontugne, Patrice Abry, Kensuke Fukuda, Pierre Borgnat, Johan Mazel, Herwig Wendt, Darryl Veitch. **Random Projection and Multiscale Wavelet Leader Based Anomaly Detection and Address Identification in Internet Traffic**. In *Proceedings of the 40th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. Brisbane, Australia, April 2015. [pdf]

12. Romain Fontugne, Johan Mazel, Kensuke Fukuda. **An Empirical Mixture Model for Large-Scale RTT Measurements**. In *Proceedings of the 34th IEEE International Conference on Computer Communications (INFOCOM 2015)*. Hong Kong, April 2015. [pdf]

13. Michał Kruczkowski, Ewa Niewiadomska-Szynkiewicz, Adam Kozakiewicz. **FP-tree and SVM for Malicious Web Campaign Detection**.

In *Proceedings of the 7th Asian Conference on Intelligent Information and Database Systems (ACIIDS 2015)*, Lecture Notes in Computer Science vol. 9012, 193-201, Bali, Indonesia, March 2015.

14. Rishikesh Sahay, Gregory Blanc, Zonghua Zhang, Hervé Debar. **Towards Autonomic DDoS Mitigation using Software Defined Networking**. In *Proceedings of the 2015 NDSS Workshop on Security of Emerging Networking (SENT 2015)*. San Diego, CA, US, February 2015. [pdf]

15. Michalis Athanasakis, Elias Athanasopoulos, Michalis Polychronakis, Georgios Portokalidis, and Sotiris Ioannidis. **The Devil is in the Constants: Bypassing Defenses in Browser JIT Engines**. In *Proceedings of the 22nd Network and Distributed System Security Symposium (NDSS 2015)*. San Diego, CA, US, February 2015. [pdf]

16. Jianxing Chen, Romain Fontugne, Akira Kato, and Kensuke Fukuda. **Clustering Spam Campaigns with Fuzzy Hashing**. In *Proceedings of the 10th Asian Internet Engineering Conference (AINTEC'14)*. Chiang Mai, Thailand, November 2014. [pdf]

17. Sirikarn Pukkawanna, Youki Kadobayashi, Gregory Blanc, Joaquin Garcia-Alfaro, Hervé Debar. **Classification of SSL Servers based on their SSL Handshake for Automated Security Assessment**. In *Proceedings of International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS 2014)*. Wroclaw, Poland, September 2014. [pdf]

18. Hajime Tazaki, Kazuya Okada, Yuji Sekiya, Youki Kadobayashi. **MATATABI: Multi-layer Threat Analysis Platform with Hadoop**. In *Proceedings of International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS 2014)*. Wroclaw, Poland, September 2014. [pdf]

19. Daisuke Miyamoto, Takuji Iimura, Gregory Blanc, Hajime Tazaki, Youki Kadobayashi. **EyeBit: Eye-Tracking Approach for Enforcing Phishing Prevention Habits**. In *Proceedings of International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS 2014)*. Wroclaw, Poland, September 2014. [pdf]

20. Jean Lorchat, Cristel Pelsser, Romain Fontugne. **Collaborative Repository for Cybersecurity Data and Threat Information**. In *Proceedings of International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS 2014)*. Wroclaw, Poland, September 2014. [pdf]

21. Michal Kruczkowski, Ewa Niewiadomska-Szynkiewicz. **Support Vector Machine for malware analysis and classification**. In *Proceedings of Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence, pp.415-420*. Warsaw, Poland. August, 2014.

22. Jun Liu, Kensuke Fukuda. **Towards a Taxonomy of Dakrnet Traffic**. In *Proceedings of the International Workshop on Traffic Analysis and Characterization (TRAC 2014)*. Nicosia, Cyprus, August 2014. [pdf]

23. Johan Mazel, Romain Fontugne, Kensuke Fukuda. **A Taxonomy of Anomalies in Backbone Network Traffic**. In *Proceedings of the International Workshop on Traffic Analysis and Characterization (TRAC 2014)*. Nicosia, Cyprus, August 2014. [pdf]

24. Martina Lindorfer, Stamatis Volanis, Alessandro Sisto, Matthias Neugschwandtner, Elias Athanasopoulos, Federico Maggi, Christian Platzer, Stefano Zanero, Sotiris Ioannidis. **AndRadar: Fast Discovery of Android Applications in Alternative Markets**. In *Proceedings of the 11th Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA)*. London, UK, July 2014. [pdf]

25. Kazuya Okada, Hiroaki Hazeyama, Youki Kadobayashi. **Oblivious DDoS Mitigation with Locator/ID Separation Protocol**. In *Proceedings of the 9th International Conference on Future Internet Technologies*. Tokyo, Japan, June 2014. [pdf]

26. Thanasis Petsas, Giannis Voyatzis, Elias Athanasopoulos, Michalis Polychronakis, Sotiris Ioannidis. **Rage Against the Virtual Machine: Hindering Dynamic Analysis of Mobile Malware**. In *Proceedings of the 7th European Workshop on Systems Security (EuroSec)*. Amsterdam, The Netherlands, April 2014. [pdf]

27. Romain Fontugne, Johan Mazel, Kensuke Fukuda. **Hashdoop: A MapReduce Framework for Network Anomaly Detection**. In *Proceedings of the 2nd International Workshop on Security and Privacy in Big Data (BigSecurity2014) in conjunction with IEEE INFOCOM2014*. Toronto, Canada, April 2014. [pdf]

28. Daisuke Miyamoto, Satoru Teramura, Masaya Nakayama. **INTERCEPT: High-interaction Server-type Honeypot based on Live Migration**. In *Proceedings of the 2nd Workshop on Emulation Tools, Methodology and Techniques (EMUTools 2014)*. Lisbon, Portugal, March 2014. [pdf]

29. Johan Mazel, Romain Fontugne, Kensuke Fukuda. **Visual comparison of Network Anomaly Detectors with Chord Diagrams**. In *Proceedings of the 29th Symposium on Applied Computing (SAC)*. Gyeongju, Korea, March 2014. [pdf]

30. Hajime Tazaki, Frederic Urbani, Emilio Mancini, Mathieu Lacage, Daniel Camara, Thierry Turletti, Walid Dabbous. **Direct Code Execution: Revisiting Library OS Architecture for Reproducible Network Experiments**. In *Proceedings of the 9th International Conference on emerging Networking EXperiments and Technologies (CoNEXT)*. Santa Barbara, California, December 2013. [pdf]

31. Antonis Papadogiannakis, Laertis Loutsis, Vassilis Papaefstathiou, Sotiris Ioannidis. **ASIST: Architectural Support for Instruction Set Randomization**. In *Proceedings of the 20th ACM Conference on Computer and Communications Security (CCS)*. Berlin, Germany, November 2013. [pdf]

32. Yuta Kazato, Kensuke Fukuda, Toshiharu Sugawara. **Towards classification of DNS erroneous queries**. In *Proceedings of the 9th Asian Internet Engineering Conference (AINTEC)*. Chiang Mai, Thailand, November 2013. [pdf]

33. Kazuya Okada, Yuji Sekiya, Youki Kadobayashi. **A Design Consideration for SDN-based Internet eXchange**. In *Proceedings of the 18th Internet Conference (IC2013)*. Tokyo, Japan, October 2013. [pdf]

2.2 Publications in journals

1. Changhoon Yoon, Taejune Park, Seungsoo Lee, Heedo Kang, Seungwon Shin, Zonghua Zhang. **Enabling security functions with SDN: A feasibility study**. In *Computer Networks*, Volume 85, Pages 19–35. July, 2015. [pdf]

2. Yuji Sekiya, Tomohiro Ishihara and Hajime Tazaki. **DNSSEC simulator for realistic estimation of deployment impacts**. In *IEICE Communications Express*, Vol.3, No.10, 305–310. October, 2014. [pdf]

3. P. Szynkiewicz, A. Kozakiewicz. **System for off-line generation of signatures of active threats (System wytwarzania off-line sygnatur zagrożeń aktywnych)**. In *Telecommunications Review and Telecommunications News (Przegląd telekomunikacyjny i wiadomości telekomunikacyjne)*, vol. 8-9, 1090–1098. September 2015 (in Polish).

4. A. Kozakiewicz, T. Pałka, P. Kijewski. **Detection of addresses of botnet C&C servers in data from sandbox environments (Wykrywanie**

adresów serwerów C&C botnetów w danych ze środowisk sandbox). In *Telecommunications Review and Telecommunications News (Przegląd telekomunikacyjny i wiadomości telekomunikacyjne)*, vol. 8-9, 1223-1231. September 2015 (in Polish).

5. M. Kruczkowski. **System for detection of malware campaigns (System do wykrywania kampanii złośliwego oprogramowania)**. In *Telecommunications Review and Telecommunications News (Przegląd telekomunikacyjny i wiadomości telekomunikacyjne)*, vol. 8-9, 789-797. September 2015 (in Polish).

6. Michal Kruczkowski, Ewa Niewiadomska-Szynkiewicz. **Comparative study of supervised learning methods for malware analysis**. In *Journal of Telecommunications and Information Technology (JTIT), Vol.4/2014, pp.1-10*, December 2014.

2.3 Deliverables

1. Deliverable **D1.4: Threat Data Final Report** (March 2016) [please add link]

2. Deliverable **D2.1: Threat Analysis** (November 2014) [pdf]

3. Deliverable **D2.2: Threat Analysis Platform** (November 2015) [pdf]

4. Deliverable **D3.1: Policy Enforcement Point Survey** (November 2013) [pdf]

5. Deliverable **D3.3: Security Information Exchange - Results** (May 2015) [pdf]

6. Deliverable **D3.5: Countermeasure Application - Results** (November 2015) [pdf]

7. Deliverable **D4.1: Requirements and Specifications of Testing Environments** (February 2015) [pdf]

8. Deliverable **D4.2: Demonstrators Evaluation** (March 2016) [please add link]

9. Deliverable **D4.3: Demonstrators Videos** (March 2016) [please add link]

10. Deliverable **D5.3: Japan Workshop Proceedings** (November 2015) [please add link]

11. Deliverable **D5.4: EU Workshop Proceedings** (October 2014) [pdf]

12. Deliverable **D5.5: Final Report on the Use and Dissemination of Knowledge** (March 2016) [please add link]

2.4 Dissertations

1. Michał Kruczkowski. **Analysis of malicious campaigns in multiple heterogeneous threat datasources.** PhD dissertation, in Polish, defended at the Systems Research Institute of the Polish Academy of Science on 2015.11.20.

2.5 Technical reports

1. Hajime Tazaki, Kazuya Okada, Yuji Sekiya, Youki Kadobayashi. **MATATABI: Multi-layer Threat Analysis Platform with Hadoop**. In *IEICE Technical Report*, Vol. 113, No. 502, IEICE-ICSS2013-77, pp.113-118, March 2014.

2. Tomohiro Ishihara, Hajime Tazaki, Kazuya Okada, Daisuke Miyamoto, Yuji Sekiya. **DNS Traffic Analysis Platform with Hadoop Framework**. In *IEICE Technical Report*, Vol. 113, No. 502, IEICE-ICSS2013-80, pp.131-135, March 2014.

2.6 Posters

1. Thanasis Petsas, Kazuya Okada, Hajime Tazaki, Gregory Blanc, Paweł Pawliński. **A Trusted Knowledge Management System for Multi-layer Threat Analysis**. In *7th International Conference on Trust & Trustworthy Computing (TRUST)*. Heraklion, Crete, June 30 July 2, 2014. [Abstract] [Poster]

2. Takuji Iimura, Daisuke Miyamoto, Hajime Tazaki, Youki Kadobayashi. **NECOMAtter: Curating Approach for Sharing Cyber Threat Information**. In *9th International Conference on Future Internet Technologies. Tokyo, Japan*. June 2014. [pdf]

3. Sirikarn Pukkawanna, Hiroaki Hazeyama, Youki Kadobayashi, Suguru Yamaguchi. **Detecting Anomalies in Massive traffic with Sketches**. In *9th International Conference on Future Internet Technologies*. Tokyo, Japan. June, 2014. [pdf]

4. **Research and Academic Computer Network (NASK)**. In *3rd Plenary Meeting*. Tokyo, Japan, January 2015.

5. Thanasis Petsas. **Rage Against the Virtual Machine: Hindering Dynamic Analysis of Mobile Malware**. In *7th European Workshop on Systems Security (EuroSec)*. Amsterdam, The Netherlands. April 2014. [Poster]

6. Institut Mines-Telecom. **On the Use of Data Mining Techniques for the Clustering of URLs Extracted from Network-based Malware Traces**. In *2nd Plenary Meeting*. Kyoto, Japan, January 2014. [Poster]

7. **Research and Academic Computer Network (NASK)**. In *2nd Plenary Meeting*. Kyoto, Japan, January 2014. [Poster]

8. University of Tokyo (UT). **NECOMATter: twitter saves the (cyber) world**. In *2nd Plenary Meeting*. Kyoto, Japan, January 2014. [Poster]

9. Hajime Tazaki (UT), Kazuya Okada (NAIST). **NECOMA Multilayer Threat Data Collection and Analysis Platform with Hadoop**. In *2nd Plenary Meeting*. Kyoto, Japan, January 2014. [Poster]

10. Hajime Tazaki (UT), Tomohiro Ishihara (UT), Yuji Sekiya (UT). **Design and Implementation of DNSSEC Simulator using Unmodified Real Implementations**. In *2nd Plenary Meeting*. Kyoto, Japan, January 2014. [Poster]

2.7 Talks and presentations

1. Youki Kadobayashi, **Cyber-resilience and International Cooperation**. At *JSPS Security Symposium*. 28th March 2014, Tokyo, Japan.

   Conference Link: https://www.ht.sfc.keio.ac.jp/jsps-sec2014/

2. Daisuke Miyamoto. **Boosting Human: Beyond Anti-Phishing Technologies**. At *5th APT Cybersecurity Forum (CSF-5)*. 27th May 2014, Ulaanbaatar, Mongolia.

   Conference Link: http://www.aptsec.org/2014-CSF5

3. Yuji Sekiya. **NECOMA Project objectives and results**. At *JP-EU Workshop on Cloud Computing Research*. 31st, July 2014, Tokyo, Japan.

   Conference Link: http://www.ocean-project.eu/bin/view/Events/EU_Japan_Cloud_Research_31_July_2014

4. Yuji Sekiya. **PIX-IE : Programmable Internet eXchange in Edo**. At *Asia-Pacific Information Infrastructure (APII) Workshop 2014*. 8th, October 2014, Osaka, Japan.

   Conference Link: http://www.jgn.nict.go.jp/english/reports/presentation/APII_ws-2014.html

5. Hajime Tazaki, Hervé Debar. **NECOMA Project Nippon-European Cyberdefense-Oriented Multilayer threat Analysis**. At *5th EU-Japan Symposium on ICT Research and Innovation*. 15th October 2014, Brussels, Belgium.

   Conference Link: http://www.ict-fire.eu/events/past-events/5th-eu-japan-symposium-in-ict-research-and-innovation.html

6. Youki Kadobayashi, **Cyber-resilience and International Cooperation: Interim Report from the NECOMA Project**. At *7th Anti-malware Engineering Workshop*. 22nd October 2014, Hokkaido, Japan.

   Conference Link: http://www.iwsec.org/mws/2014/en.html

7. Daisuke Miyamoto, **End-user Protection in Cybersecurity - Aspect from Cognitive Psychology**. At *The 2014 Cybersecurity Data Mining Competition and Workshop*. 31st, October 2014, Kuala Lumpur, Malaysia.

   Conference Link: http://www.csmining.org/cdmc2014/index.php?id=16

8. Paweł Pawliński. **Avoiding Information Overload: Automated Data Processing with n6**. At *26th Annual FIRST Conference*. Boston, USA. 23 June 2014.

   Online: http://www.necoma-project.eu/m/filer_public/43/e8/43e890c8-9bfd-43c9-9f1b-746c0feadcee/first_2014_-_pawlinski-_pawel_-_avoiding_information_overload.pdf

9. Paweł Pawliński. **Evaluating Threat Intelligence Feeds**. At *FIRST Technical Colloquium for Threat Intelligence*. Munich, Germany. 24 February 2016.

   Online: http://www.necoma-project.eu/m/filer_public/b9/da/b9dafadd-adf8-4875-afd5-2d188dd96449/pawel-pawlinski-evaluating-ti-feeds.pdf

3 Dissemination of dataset

This chapter provides the overview of our published dataset. Out of the collected threat information summarized in Deliverable 1.2 and 1.3, we have plans to publish our dataset along with the respective dissemination levels: *Public*, *Restricted to other program participants*, *Restricted to a group specified by the consortium*, and *Confidential* (only for members of the consortium).

3.1 MAWI traffic dataset (publicly available)

The MAWI traffic dataset[2] is composed of packet traces with short payloads, collected in both directions on a 150 Megabit Ethernet link, which connects WIDE and its upstream network. Data is taken on a daily basis from 14:00 JST for 15 minutes. The first 96 bytes including the Ethernet frame are captured for each packet. NTP is used for clock synchronization. Some flows are observed only in one direction due to asymmetric routing.

The dataset is available to download at http://mawi.wide.ad.jp/mawi/samplepoint-F/2014/.

As part of a Day In The Life of the Internet (DITL) project, other datasets are also made available at http://mawi.wide.ad.jp/mawi/ditl/ditl2013/ for a 72-hour-long trace dataset from 2013, http://mawi.wide.ad.jp/mawi/ditl/ditl2014/ for a 24-hour-long trace dataset from 2014, http://mawi.wide.ad.jp/mawi/ditl/ditl201412/ for another 24-hour-long trace dataset from 2014, and http://mawi.wide.ad.jp/mawi/ditl/ditl2015/ for a 48-hour-long trace dataset from 2015.

3.2 Agurim multi-demensional aggregated flow traffic dataset (publicly available)

This sample dataset is an effort to promote measurement data sharing and provide broader access to backbone traffic for the benefit of the networking community.

The data has been sampled from the transit link of the WIDE network (AS2500) in Japan since February 2013. IP addresses appearing in the dataset are anonymized using a prefix-preserving method. More information about the dataset is available from the Agurim traffic dataset site[1].

The main view of the Agurim Web interface provides dual plots, a volume-based plot on the left and a packet-based plot on the right. Each plot presents 7 significant aggregate flows by default. The legend label shows each aggregate flow with the main attribute and its share against the total traffic, along with the sub-attributes and their shares within the aggregate flow. In the address view, the main attribute is source and destination addresses and the sub-attributes are protocols. In the protocol view, the main attribute is protocol and the sub-attributes are addresses. Addresses are presented with thier prefix length when aggregated. '*' is a wildcard (e.g., 0.0.0.0/0 for IPv4 address) but '*::' is used for IPv6 address. Protocols are presented by proto:sport:dport (e.g., '6:80:1234' for proto=TCP, sport=80, dport=1234).

3.3 FORTH's dataset (local NECOMA database available through n6, currently available to partners)

The *NECOMA* dataset hosted at FORTH contains information gathered from a variety of sensors. These include the following: FORTH honeypots data (AMUN [2] deployment) which contain information on cyber-attacks gathered by monitoring an unused network address space, data captured by Dionaea [3], another low-interaction honeypot deployment that captures attack payloads and malware, as well as data gathered from a set of publicly accessible web sources. Below are the sensors, from which data are currently collected and are available through n6:

- BladeDefender http://www.blade-defender.org/
- PhishTank http://www.phishtank.com/
- SANS http://www.sans.org/

[1] https://mawi.wide.ad.jp/~agurim/
[2] Amun: Python Honeypot http://amunhoney.sourceforge.net/
[3] Dionaea honeypot: http://dionaea.carnivore.it/

- OffensiveComputing http://www.offensivecomputing.net/
- Threat Expert http://www.threatexpert.com
- MD:PRO http://frame4.net/

3.4 UT's human behavioural dataset (restricted to a group specified by the consortium)

This dataset contains two types of data: one is concerned with decision results and criterion collected by means of questionnaire, while the other is concerned with eye movement records collected by an eye-tracking camera.

From December 2013 to February 2014, we performed a participant-based experiment to precisely measure end-user's eye activities. It should be noted that our experiments do not collect and/or analyze personally identifiable information. Our experiment displayed 20 website screenshots, including legitimate websites and pseudo phishing sites, invited participants to judge whether the site seems to be phishing or not, and asked their criteria while assessing website's credibility. Additionally, we also performed a set of experiments in August 2014, and from December 2015 to February 2016.

The eye tracking data is composed of two types of files. One is a video file, in AVI format, that captures eye position and eye movements on the screen, and the other one is a CSV file containing records of eye movement, namely time, eye position on the screen, and category of eye movements, e.g., saccadic, fixation, etc. The amount of dataset is 10GBytes, and the data involves 70 participants' eye tracking data.

3.5 IMT's SSL dataset

The dataset is comprised of response payloads from stimulated SSL servers covering the whole IPv4 range. These responses include information on the responding server IP address, the SSL/TLS protocol version used, and the ciphersuites negotiated, as well as the server certificate contents.

The SSL/TLS measurement campaigns were first launched in July 2010 and several times during the year 2011, prior to the *NECOMA* project. They were part of an experiment to assess the quality of HTTPS servers [8]. During the months of March and April 2014, another campaign was launched to refresh the information contained in these datasets. The provided datasets include these campaigns, as well as, some other measurement campaigns, the results of which are publicly available and released by other institutions such as the Electronic Frontier Foundation [4], amounting to a total of 21 campaigns available.

The collection campaigns performed at Télécom SudParis were based on active enumeration of open HTTPS ports (TCP-/443) over the entire IPv4 space. The dataset collected in July 2010 was initiated through a single TLS `ClientHello` message directed to servers that responded on port 443, while the 7 campaigns collected in July 2011 featured several different `ClientHello` messages with varying protocol versions, ciphersuites, and TLS extensions. The final 11 campaigns collected during March/April 2014 were featuring even more diversity in the parameters used to build the requests.

The dataset is accessible at: `http://phoenix.telecom-sudparis.eu:2534/ssl_n6/<command>.json?<parameters>`. A list of available commands, and details on the syntax are availabe in Deliverables D1.4 and D3.3, respectively.

3.6 n6 platform provided by NASK

The n6 (Network Security Incident eXchange) platform[4] was created by NASK[5] for the purpose of systematic collection, processing, and sharing of cyber-security information. Most of the data available in the platform can be characterized as "detection indicators," [12] i.e. technical, machine-readable information on various threats (e.g. known attack sources, compromised machines, vulnerable services) that can be used for the purpose of defense and remediation. In 2015 n6 was used to process approximately one billion security events.

Data sources integrated into the platform include internal systems run by CERT Polska (part of these were described in deliverables D1.2, D1.3, and D1.4), however the majority of them are provided by third-party organizations, including technology companies, non-profit organizations, and independent researchers. At the time of writing (March 2016), n6 is collecting data from over 70 different sources.

Access to the platform is offered as a free service for the community. Any organization in Poland, including, but not limited to, ISPs, industry, and government institutions, can subscribe to it and receive all available data relevant to its network. Currently there are over 200 subscribed organizations in Poland, from a wide range of sectors.

Information from sources maintained by CERT Polska itself is also shared with 25 national and governmental CERTs in Europe and worldwide. By producing actionable information and sharing it with responsible entities NASK can contribute to the improvement the overall security of the internet ecosystem. Finally, n6 is used to share data for research purposes internally within NASK and with the NECOMA consortium partners.

[4] Website: `http://n6.cert.pl/` (Polish)

[5] CERT Polska, a part of NASK, is the CERT with the national responsibility in Poland.

The platform is geared to automated, high-volume, and low-latency information sharing. In the course of the NECOMA project two complementary application interfaces (APIs) were developed: REST API (using HTTP and multiple data formats including JSON and IODEF [14]) and stream API (based on STOMP [1] and JSON). These interfaces were described in detail in deliverables D1.1 and D3.2.

4 Dissemination of published tools

This chapter describes our contributions on tool development that are publicly disseminated in order to deliver our efforts to others. These contributions help the activity of the NECOMA project in some way. The terminology "tool" refers to software used/developed for our proposed system, or internally used for research purposes like data analysis programs. Moreover, "tools" also includes any services such as web based applications. Part of our developed software is published using GitHub[1].

4.1 Modified tools for MATATABI

MATATABI is a threat analysis system based on Apache Hadoop software [16]. Some additional tools internally used by MATATABI are modified and updated in order to meet our usage.

hadoop-pcap

hadoop-pcap is a library originally developed by Réseaux IP Européens Network Coordination Centre (RIPE NCC). It offers a seamless interface for accessing `pcap` data stored in Hadoop Distributed Filesystem (HDFS). Our analysis heavily depends on the `pcap` dataset but a couple of issues in hadoop-pcap library prevents us to use it for analyzing our `pcap` files.

All the effort on the development of this library is about fixing issues in order to make it work with our samples.

presto-db

Presdo-db (or Presto) is a distributed SQL query engine developed by Facebook. It is a framework working with a couple of data sources including HDFS and traditional database engines, such as MySQL and Postgresql. Since presto-db outperforms better than Apache Hadoop software in a par-

[1] https://github.com/necoma/

ticular case [16], we have used presto-db as a query interface. The detailed benchmark results are shown in our BADGERS 2014 paper [16].

Since presto-db was a young software and not mature enough when we started to use it, we needed to modify the original software, which is publicly available. Our development regarding presto-db was only to port a currently unsupported platform so that it can be deployed with MATATABI.

4.2 NECOMAtter

NECOMAtter is a system intended to ensure dissemination and spread of threat information. For this reason, it facilitates collaboration among information providers and users to bring together various stakeholders with different expertise. Thus, we designed NECOMAtter for accessing large amount of diverse data, and expediting development of improvised defenses against improvised attacks through ad-hoc collaboration of NECOMAtter users.

We are also inspired by the concept of timeline, a real-time list of tweets on Twitter. In the case of NECOMAtter, there are also various types of cyber threat information, and each NECOMAtter user have different demands for this information. For helping the users to retrieve their suitable information along with the demands, NECOMAtter is equipped with the timeline feature. It not only allows them to search information by tags, but also enables to select by subscribing other NECOMAtter users and/or curated groups of the users.

4.3 Eye-bit

EyeBit is a system designed to induce end users habit of checking the address bar of the web browser. To detect phishing, users should check a website's URL and an SSL padlock icon displayed at the address bar. However, users often ignore the address bar under the phishing attack. We assume that the habit of checking the bar will improve users' awareness of secure information.

EyeBit employs an eye-tracking device to check whether a user can see the address bar or not. At first, it deactivates all input forms. When it detects that the user has checked the browser's address bar, all input forms are then activated. Eye-bit is developed as a browser extension (**eyebit_chrome_extension**). Another component of the system is the **eyebit_server**. **eyebit_server** interacts to the extension and the eye-tracking camera, e.g., EyeTribe Tracker [2]. The EyeBit was described in detain in [11].

[2] https://theeyetribe.com/

4.4 WebVisor

WebVisor is a web service that provides a clustering analysis and signature generation framework to analyze a dataset of HTTP requests in a privacy-preserving manner. Registered users submit their dataset to WebVisor which will immediately strip the URLs from their domain information. Since WebVisor focuses on the path and the query string, it will not affect the analysis.

The clustering framework actually relies on a 2-step workflow, with a first statistical clustering performed over character occurrence statistics, while a second stage further refines clusters using density-based clustering. The second step is actually computed over abstracted key-value pairs representing the query string so that similar query strings in structure are clustered together, regardless of values of the parameters. The user has the possibility to tweak some parameters such as the algorithms used for two-step clustering: clustering algorithms, distance functions, number of centroids, distance between points, etc. Additional functionality includes cluster quality assessment, cluster visualization and a module to generate signatures with the limit of one signature per cluster.

A demonstration of the tools is available at: http://j.mp/WVProto.

4.5 DNSSEC simulation for ns-3

DNSSEC is an important technology designed to protect user's name query, which is an indispensable process in Internet. However, DNSSEC is not widely deployed now due to side effects in operator's point of view. Thus, operators require an evaluation framework of such a big deployment in advance, but 1) evaluating the impact on DNSSEC deployment in a practical scale is not easy task and 2) evaluation at scale with abstracted model does not either reflect realistic parameters such as delays generated by computations.

Our developed DNSSEC simulator aims to solve two difficult issues: it can estimate and simulate the impacts using actual DNS queries, DNS topology, and actual DNS implementations. Moreover, there is no need of knowledge of DNS simulation. The simulator estimates the impacts easily by providing DNS query log. The software can contribute to the deployment of DNSSEC and achieve a safe DNS service.

All the implementation and documents including instructions are publicly available[3]. More information is available on the web site and our paper [15].

createzones

Createzones is a software which constructs pseudo DNS tree for DNS emulation/simulation. In many cases, a simulation is carried out with real

[3] http://dnssec.sekiya-lab.info/

monitored DNS traffic data. For that purpose, this script creates DNS configuration files and zone files from DNS traffic data. Hence, the reproduced DNS tree by the tool can respond to the monitored DNS traffic.

Createzones script works as follows:

1. Pick up zone names from BIND9 log file or DNS traffic data. If name list contains "www.example.com" line, this module cosiders there are two zones: "example.com" and "com".
2. Assign simulation nodes as authoritative server to each zones.
3. Create zone files according to the result of step 1. Each zones contain randomly generated hostnames.
4. Create DNSSEC Key-Signing-Key (KSK) and Zone-Signing-Key (ZSK) for each zone. A signature for KSK (DS Record) should be signed by parent zone's ZSK to establish trusted chains. Therefore add the DS Record to its parent's zone file.
5. Sign each zone by their own KSK/ZSK.

Createzones script is published on GitHub [4].

4.6 The Agurim multi-dimensional flow aggregation tools

Agurim is a network traffic monitor based on flexible multi-dimensional flow aggregation in order to identify significant aggregate flows in traffic. A user can dynamically switch views based on traffic volume or packet counts, address or protocol attributes, with different temporal and spacial granularities. The supported data sources are pcap, sFlow, and netFlow.

Agurim employs a two-stage multi-dimensional flow aggregation scheme. The primary aggregation stage is for efficiently processing a huge volume of raw traffic records (e.g., pcap or NetFlow). It reads raw traffic records and then produces initial aggregated flow records. The secondary aggregation stage is for providing flexible views to users. It reads the output of the primary aggregation stage, and produces concise summaries as specified by the aggregation parameters in a user request.

We have developed a set of tools, named 'Agurim', as open source software. The primary aggregation tool works as a collector of pcap, NetFlow or sFlow, and generates initial aggregated flow records. The secondary aggregation tool implements a more elaborate aggregation engine with several control parameters. The Web user interface allows a user to dynamically

[4] https://github.com/shored/createzones/

switch views based on traffic volume or packet counts, address or protocol attributes, with different temporal and spatial granularity.

Furthermore, we have made anonymized open dataset collected from the WIDE backbone network to demonstrate the feasibility and usefulness of multi-dimensional flow aggregation. It allows researchers and operators to browse real traffic through the Agurim Web user interface, which has a great potential for advanced security research.

The source code and related documents are available on the Agurim site[5].

4.7 The Tamias distributed storage system

The Tamias Project[9] is an attempt to create an open-source, privacy-aware distributed file storage in order to provide:

- Secure and reliable storage for all kind of files
- Easy sharing with per-object control
- Storage provider independence

Tamias is based on the Tahoe-LAFS storage system[18] and borrows its secure properties while leveraging a public-key infrastructure in order to provide distributed access control.

More details of the Tamias Distributed Storage System and how it is used in the *NECOMA* project can be found in deliverable, *Security information exchange design* (D3.2).

The source code and related documents are available on the Tamias project site[6].

4.8 n6 SDK

n6 is a system for exchange and processing of security-related information created by CERT Polska[7]. The initial goals of the system were to improve management of multiple data sources and effective distribution of data to constituents and partner security organizations. One of its main features is the REST API, which provides a unified way to share many types of security information and different kinds of indicators.

The n6 REST API is one of the core components of the NECOMA platform, since it is the main machine-to-machine communication mechanism

[5] https://mawi.wide.ad.jp/~agurim/about.html
[6] https://tamias.iijlab.net/
[7] CERT Polska, a part of NASK, is the CERT with the national responsibility in Poland.

between different components. The overall architecture is described in deliverable D2.1 and details of the information exchange are described in deliverables D1.1 and D3.2.

The n6 SDK (Software Development Kit) is a software package that allows to add an n6-compatible API to any data store with little effort. The SDK is used by the members of the NECOMA consortium to exchange data, both during initial collection (data from honeypots, sinkholes, etc.), and for providing information to the resilience mechanisms. An extensive SDK tutorial for developers is included in deliverable D3.2.

The n6 SDK was released on an open source licence (GPL) in December 2014. The source code is available on GitHub: https://github.com/CERT-Polska/n6sdk. n6 SDK is under active development which will continue beyond the end of NECOMA.

4.9 Hashdoop

Hashdoop [5] is a MapReduce framework for anomaly detection in Internet backbone traffic. The goal of Hashdoop is to leverage the scalability, scheduling and fault tolerance benefits of Hadoop clusters to efficiently analyze Internet backbone traffic. We found that the classical Hadoop data slicing used for textual documents breaks spatial and temporal traffic structures, which dramatically deteriorates anomaly detectors performance. Hashdoop preserves spatial and temporal traffic structures by splitting the traffic with a hash function, thus permitting anomaly detectors to be processed with the MapReduce model.

The source code of Hashdoop is publicly available on Github https://github.com/necoma/hashdoop.

4.10 DNS DDoS defense and countermeasure

DNS DDoS Defense and Countermeasure (d4c) is a deep packet inspection and firewall system for DNS based DDoS attack. d4c works as a layer-2 inline device with two interfaces. Packets from an interface are forwarded to another interface. When forwarding packets, d4c filters invalid DNS packets. In addition, d4c uses netmap for its data-plane for achieving high throughput. d4c has two functions for packet forwarding and filtering: 1) filtering DNS responses from rogue resolvers in outer networks and 2) filtering DNS queries that have specified QNAME.

DNS based DDoS traffic is usually generated by open resolvers around the world. Assuming that correct DNS responses are sent from correct resolvers, DNS responses from resolvers in outer networks are fully incorrect and they are DDoS traffic. d4c protects a network from DNS based DDoS

traffic by filtering this incorrect DNS response. Response filtering example is shown below.

```
d4c -l netmap:p2p1 -r netmap:p2p2 -d 10.10.0.0/16 -d 172.16.0.0/16
-s 8.8.8.8/32
```

`-l` and `-r` options indicate two interfaces. Outer networks that rogue resolvers located in and correct resolvers are configured by `-d` and `-s` options. If the destination IP address of response packet is included in the prefixes specified by `-d` options and the source IP address of the packet is included in the prefixes specified by `-s` options, the packet is blocked.

The next example describes a function for blocking a cause of DNS based DDoS attacks. DNS based DDoS traffic is generated by invalid DNS queries from infected or malicious clients. These DNS queries usually contain the same or similar QNAMEs. To block these queries in order to reduce DDoS traffic, d4c has QNAME suffix match blocking. Blocking QNAME is configured by `-m` option.

```
d4c -l netmap:p2p1 -r netmap:p2p2 -m hoge.com -m huga.com
```

In this case, queries for `.*hoge.com` and `.*huga.com` are droped.

The source code of d4c is pulicly available on Github https://github.com/upa/d4c.

5 Dissemination through standards activities

This chapter briefly summarizes our activities on international standardization. Towards deployment of our technologies to the Internet, we have been participating in the standards activities of Internet Engineering Task Force (IETF) and International Telecommunication Union Telecommunication Standardization Sector (ITU-T). Our motivation is to maximize the chance for sharing our tools, knowledge and methodologies developed in the *NECOMA* project.

5.1 Internet Engineering Task Force

In the NECOMA project, we decided to use n6 for exchanging cyber threat information. Basically, n6 uses an event-based data model for representation of all types of security information. Each event is natively represented as a JSON object with a set of mandataory and optional attributes.

For the successful dissemination of our activities, we have engaged in international standards meetings. In the Internet Engineering Task Force (IETF), the Managed Incident Lightweight Exchange (MILE) working group develops standards to support computer and network security incident management.

Our internet drafts presented at IETF are as follows.

5.1.1 Knowledge obtained from the implementation experience of an IODEF-capable incident response management system

This draft explains our observation on the usability of IODEF [3], based on our experiments. We aim at developing an IODEF-capable incident response management system in order to facilitate incident response activities. We

started to design and implement the system for CERT. However, there are several technical issues which arose while implementing and operating the system. The draft shares the observation from our prototype implementation.

In our implementation, we encountered problems while dealing with XML schema. To reduce the development cost, we employed code generators that build class libraries for accessing values in IODEF elements. Due to the complexity of IODEF message format in RFC5070, some code generators could not understand this schema. We also found some operational problems as well as implementation problems. Most of the problems were on the choice of values for IODEF attributes and/or elemens.

Implementation Issues Since a code generator for XSD automatically develops useful libraries for accessing XML attributes and/or composing messages, we tried to build the libraries from RFC 5070. However, some generators could not properly understand the schema due to the complexity of IODEF XSD.

An alternative problem is that IODEF uses '-' (hyphen) symbols in its classes or attributes. According to the language specification, it usually does not contain '-' symbols in the class and/or function names, and hence, some generators could not build the libraries.

Operational Issues In some cases, we were not sure to choose the suitable attributes for incidents. For example, Incident Type class is used to categorize incidents into several types. However, there already exist various incident classifications, and it is often hard to fit them into the type attribute of the impact class.

The numbering of Incident ID needs to be considered. Otherwise, information, such as the number of incidents within certain period could be observed by document receivers.

5.1.2 MILE implementation report

This document is a collection of implementation reports from vendors, consortiums, and researchers who have implemented one or more of the standards published from the IETF INCident Handling (INCH) and Management Incident Lightweight Exchange (MILE) working groups.

This draft explains the n6 since it supports alternative output data formats for keeping compatibility with existing systems - IODEF and CSV.

5.2 International Telecommunication Union Telecommunication Standardization Sector

In the International Telecommunication Union Telecommunication Standardization Sector (ITU-T), Study Group 17 (SG17) deals with a broad range of standardization issues on security. In particular, Question 4 (Q.4) Cybersecurity group defined X.1500 series to exchange threat information. Our platform, n6 was also presented at the July 2014 interim meeting of Q.4. NECOMA also proposed the following two work items that we considered in Task 3.4.

5.2.1 X.cogent, Design considerations for improved end-user perception of trustworthiness indicators

This draft Recommendation has been established within ITU-T SG17/Q.4 as a result of the deliberation of contributions from *NECOMA*.

Diverse kinds of attacks employ replicated contents from trustworthy service providers, thereby deceiving end-users into believing its false trustworthiness. This draft Recommendation describes design considerations for improved end-user perception of trustworthiness indicators. The appendix describes representative techniques for measuring end-user perception of such indicators.

This draft defines trustworthiness indicator as the symbols presented by a web user agent that will be used to inform the trustworthiness of the website to end users. Techniques for improved end-user perception of trustworthiness indicators include visual elements, narrative elements, peripheral design transitions, training mode and accessibility.

- Visual elements: Developers of trustworthiness indicators must consider the use of standardized visual elements. Past studies have revealed that symbolic encoding of trustworthiness indicators, e.g., in Uniform Resource Locators, are not friendly to novice users and they are often ignored [11]. It is thus recommended to introduce visual elements, e.g., icons that represent trustworthiness indication. Implementers may consider employing a few standardized visual elements, as in road signs, to minimize cognitive overhead and training overhead.

- Narrative elements: Additionally, it is recommended to equip end-user software with the capability to convert symbolic information into narrative elements that do not employ acronyms. It can also be helpful to visually impaired users, when combined with text to speech systems.

- Peripheral design: Developers of trustworthiness indicators may test their interface regarding the peripheral design transitions. Sudden

transition in peripheral vision may be effective to signal potential risk. It is thus recommended to employ this technique through transition of peripheral designs, whenever end-users are faced with high-risk websites or e-mail messages.

- Training mode: The end-user perception of risk will be inaccurate at best if he or she is very rarely exposed to such risks. It is therefore recommended to equip end-user software with training mode, where emulated risk events can be artificially generated and end-user's perception accuracy can be trained. Such training can also be incentivized by gamifying the training.

- Accessibility: Developers of trustworthiness indicators should design its interface considering the accessibility. Vision refers to the ability to distinguish the form, size, shape and color of visual stimuli. For individuals with vision impairment, there can be difficulties to find trustworthiness indicators. Due to the effects known as "protanopia" and "deuteranopia", some end-users have problems in distinguishing colors, e.g., red from green. The Telecommunications Accessibility Checklist [7] intends to ensure that the specified services and features are usable by diverse users, including people with disabilities. In order to provide better accessibility for visual impairment or blindness, the interface should provide media presentation to the user, and allow to control in various modes and types of control action.

  Screen reader applications may retrieve trustworthiness indicators from websites. They may present security information, e.g., the green address bar of an EV-SSL certificate, and read the information with text-to-speech services. They may also summarize information from Document Object Model (DOM) tree within the browser.

5.2.2 X.metric, Metrics for evaluating threat and resilience in cyberspace

This draft Recommendation has also been established within ITU-T SG17/Q.4, cybersecurity, as a result of *NECOMA* contribution.

The lack of standards in quantitatively characterizing threats, and subsequently, diverse ad-hoc attempts based on easily measurable quantities, are generating great confusion among stakeholders, sometimes referred to as port-scan counting syndrome. Failure to estimate the scale of threats in a timely manner leads to the late deployment of resilience mechanisms or the incorrect choice of resilience mechanism, subsequently diminishing its effectiveness. This draft Recommendation on Metrics for Evaluating Threat and Resilience in Cyberspace aims to study possible quantification methods for threats and associated resilience mechanisms.

6 Dissemination through collaborative activities

6.1 EU and Japan plenary meetings

The consortium have engaged in eight plenary meetings:

- September 2013 in Paris, France,
- January 2014 in Kyoto, Japan,
- July 2014 in Paris, France,
- January 2015 in Tokyo, Japan,
- July 2015 in Paris, France,
- November 2015 in Kyoto, Japan,
- January 2016 in Madrid, Spain,
- March 2016 in Heraklion, Greece.

These plenary meetings contributed to internal dissemination of knowledge and improved visibility of research and development activities among partners.

Through detailed presentations followed by interactive discussions at the plenary meetings, a good understanding has been established on available assets and ongoing activities among *NECOMA* partners. These include, but are not limited to: application programming interfaces for threat information exchange, big-data platform for cyber-threat data analytics, software tools for threat data collection, network introspection techniques, algorithms and their implementations for cyber-threat data analysis, and a variety of heuristics and measurement apparatus for end-user protection.

The consortia disseminated knowledge among partners, targeting four demonstrators in mind: DDoS mitigation, botnet introspection, malware

campaign mitigation, and smartphone user protection. This resulted in the exchange of researchers and students, joint papers, convergence to common API, increased access to datasets between EU and Japanese partners, and extended network of collaboration.

The plenary meetings also contributed to the exploitation planning of the research findings and the developed tools. For instance, the n6 API has been introduced in the standards groups; the big-data approach of *NECOMA* has been mentioned in the ENISA guideline on threat information sharing.

6.2 EU workshop (BADGERS 2014)

The 3rd International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS 2014) was co-located with the ESORICS conference in Wroclaw, Poland. The workshop was held on September 11th, 2014 and served as the confluence of computer security and general-purpose large-scale data processing. The workshop brought together people (e.g., researchers, practitioners, system administrators, security analysts) active in the emerging domain of security-related data collection and analysis for Internet-scale computer systems and networks.

The workshop included 4 papers from *NECOMA*, highlighting ongoing work in the project on the big-data platform, secure and distributed information storage, end-user protection and threat data analysis. The workshop has also served as an important means to establish touchpoints with invited speakers from both NICT and Eurecom, who shared their experience and insight on large-scale data analysis in their respective fields: network security and system security.

6.3 Japan workshop (BADGERS 2015)

The BADGERS 2015 workshop was held on November 5th, 2015, co-located with RAID conference in Kyoto, Japan, and served as a venue for research on big data for security.

The workshop included 5 papers from *NECOMA*. In these papers, we presented a framework for network traffic data collection, an automated rating algorithm for cyber threats with natural language processing, a system for analyzing visual elements to mitigate users' excessive trust in SSL/TLS, a cyberdefense method based on attack traffic redirection to decoy servers, as well as a modular framework for social forensics.

Through this workshop, participants from the Japanese security communities, participants of RAID conference, and *NECOMA* consortium members assembled to discuss the state of the art in cyber-resilience and reconfiguration mechanisms.

6.4 Teleconferences

Table 6.1: Teleconferences held for *NECOMA* project

WP	Date	Dur.	Attendees	Summary of discussions
WP1	27/09/2013	1h	NAIST, IIJ-II, NII, UT, KEIO, IMT, ATOS, FORTH, NASK, 6CURE	Report on datasets providers; Planning on the development of a common API; Agreement on the format and transport of data
WP2	03/10/2013	1h	NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress report on T2.1; Identification of results that could be realistically demonstrated by end of Y1; Discussion on general structure of D2.1; Workplan for Q4 2013
WP3	18/10/2013	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, 6CURE	Writing assignments and schedule for D3.1
WP3	05/11/2013	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, 6CURE	Work allocation review for T3.3 and T3.4; Progress report on D3.1
WP2	13/11/2013	1h	NAIST, NII, UT, IMT, ATOS, FORTH, NASK	Progress report on T2.1; Discussion on solutions to avoid duplicate work
WP3	19/11/2013	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress report on D3.1
WP1 WP2 WP3 WP5	21/11/2013	1h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress report over all active tasks (WP1, WP2 and WP3) towards Kyoto plenary meeting: Planning for Kyoto plenar y meeting
WP1	26/11/2013	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress report on T1.1 (data collection); Planning for Kyoto plenary meeting

WP2	13/12/2013	1h	NAIST, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Review of collected information for work overlap avoidance; General progress report on T2.1
WP1	13/01/2014	1h	NAIST, IIJ-II, UT, KEIO, IMT, ATOS, FORTH, NASK, 6CURE	Discussion on the relations between collected data in WP1 and their use in the foreseen use cases in WP4; Subtask planning for D1.1; Dataset description for D1.2 and D1.3 (template); Planning for Kyoto plenary meeting
WP1	17/02/2014	1h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress report on D1.1 and scheduling of tasks towards delivery; Venue discussion for a NECOMA workshop to be collocated (BADGERS)
WP3	21/02/2014	1h	NAIST, UT, IMT, 6CURE	Position report of each partner on T3.3.1 and T3.3.2; Proposal of a survey of existing DDoS mitigation schemes
WP2	24/02/2014	1h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Discussion about common tokens to cross-correlate data and analysis results consequently to Kyoto plenary meeting; Progress report on T2.1; Identfication of potential collaboration areas
WP6	24/02/2014	1h	NAIST, IMT	Discussion about intership plan (student exchange)
WP1	03/03/2014	1h	NAIST, IIJ-II, NII, UT, KEIO, IMT, ATOS, NASK, 6CURE	Review of current progress for D1.1, D1.2 and D1.3; Designation of editors and reviewers for the said deliverables

WP3	07/03/2014	1h	NAIST, UT, IMT, 6CURE	Progress report about the survey on DDoS mitigation schemes and SDN (T3.3.1/T3.3.2); Intership plan between IMT and NAIST
WP2	18/03/2014	1h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Discussion and agreement over NECOMA architecture
WP2	31/03/2014	1h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Discussion on how to manage interactions between analysis modules; Agreement on common tokens, Discussion on the architecture of the analysis platform
WP3	04/04/2014	1h	NAIST, UT, IMT, 6CURE	Discussion about DDoS mitigation threat models and autonomic computing in SDNs (T3.3.1/T3.3.2); Venue decision for joint papers
WP3	07/04/2014	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Discussion to leverage data from WP1 into D3.2; Discussions on T3.3 and T3.4 based requirements analysis for D3.2; Progress review on T3.2, T3.3 and T3.4 for M11; Planning towards M15
WP3	13/05/2014	1h	NAIST, IIJ-II, UT, IMT, 6CURE	Discussion on the survey of DDoS mitigation techniques and implementations of SDN-based techniques; Considerations about the publication of the survey to a wider audience

WP2	19/05/2014	1h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress report on analysis modules' design and prototyping; Discussion on prospective analysis modules' demonstrations at 1st annual review
WP3	20/05/2014	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress report on early work for T3.3 and T3.4
WP2	04/06/2014	1h	NAIST, IIJ-II, NII, UT, IMT, FORTH, NASK, 6CURE	Progress review concerning analysis modules; Preparation for annual review; ToC for D2.1
WP2	13/06/2014	1h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Pre-review meeting: review of planned presentations, demonstrations and scheduling
WP3	16/06/2014	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress review on T3.2 and T3.3; Pre-review discussion: planned presentations and scheduling
WP2	23/06/2014	1h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Pre-review meeting: review of demonstrations and technical presentations
WP1 WP2	23/07/2014	2h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Post-review meeting: review report discussion; T2.1 and T2.2 progress report; Discussion on work allocation for D2.1
WP3	29/07/2014	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress review on T3.2, T3.3 and T3.4; Publication of n6 SDK to partners
WP3	02/10/2014	2h	NAIST, NII, UT, KEIO, IMT, ATOS, FORTH, NASK, 6CURE	Progress review on T3.2, T3.3 and T3.4; Discussions on requirements for information exchange; ToC and work allocation for D3.2 and D3.4

WP4	14/10/2014	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress review on T4.1; Discussion on evaluation metrics; Early planning towards D4.1
WP3	27/10/2014	2h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Deliverable progress review on D3.2 and D3.4
WP2	28/10/2014	2h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Deliverable progress review on D2.1
WP5	05/11/2014	2h	NAIST, UT, IMT, ATOS, FORTH, NASK, 6CURE	Deliverable 5.1 focus meeting: review of contents, ToC and work allocation
WP2	07/11/2014	1h	NAIST, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	D2.1 progress review
WP3	12/11/2014	1h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	D3.2 and D3.4 progress review
WP2	19/11/2014	2h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	D2.1 progress review and editorial discussion
WP5	21/11/2014	2h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	D5.2 progress review and editorial discussion
WP4	03/12/2014	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress review on T4.1; Review of proposed ToC for D4.1
WP4	15/01/2015	1h	NAIST, UT, IMT	Tokyo Plenary preparation for WP4 tasks
WP4	04/02/2015	1h	NAIST, UT, IMT, ATOS, FORTH, NASK, 6CURE	Deliverable progress review on D4.1

WP4	19/02/2015	2h	NAIST, UT, IMT, ATOS, FORTH, NASK, 6CURE	Deliverable progress review on D4.1 and editorial discussion
WP3	27/02/2015	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress review on T3.2, T3.3 and T3.4; Proposed ToC draft for D3.3 and discussion
WP3	25/03/2015	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress review on T3.3 and T3.4; Discussion on D3.3 contributions
WP6	27/04/2015	2h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Y2 review preparation; Discussion on Y2 contributions for all technical WPs; Action plan for Y2 management report
WP3	07/05/2015	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, 6CURE	Deliverable progress review on D3.3; Discussion WP3 contributions to Y2 review
WP2	08/05/2015	1h	NAIST, IIJ-II, UT, ATOS, FORTH, NASK, 6CURE	Early ToC draft for D2.2; Discussion on WP2 contributions to Y2 review
WP2	29/05/2015	2h	NAIST, IIJ-II, NII, UT, IMT, ATOS, NASK, 6CURE	Selection of WP2 contributions to Y2 review
WP3	02/06/2015	1h	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Feedback on D3.3 contents; Selection of WP3 contributions to Y2 review
WP4	10/09/2015	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Discussion on feedback from Y2 review; Progress report on T4.2
WP3	16/09/2015	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Discussion on feedback from Y2 review; Progress report on T3.3 and T3.4; ToC and task assignment of D3.5

WP2	30/09/2015	1h30	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Discussion on feedback from Y2 review; ToC and task assignment of D2.2
WP2, WP4	07/10/2015	1h30	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Discussion on supplementary research to be added to D2.2; Discussion on threat metrics for WP2; Discussion on demonstrator's assessment metrics for WP4
WP3	19/10/2015	1h	NAIST, IIJ-II, UT, IMT, ATOS, FORTH, NASK, 6CURE	Progress report on D3.5; Roadmap to D3.5 delivery
WP2	23/10/2015	1h30	NAIST, IIJ-II, NII, UT, IMT, ATOS, FORTH, NASK, 6CURE	Discussion on threat metrics for WP2; Discussion on D2.2 ToC
WP4	10/11/2015	1h	IMT, 6CURE	Discussion on scenario metrics for DDoS mitigation scenario
All WPs	02/02/2016	2h	NAIST, IIJ-II, UT, IMT, ATOS, NASK, 6CURE	All WPs deliverables progress meeting and Y3 review preparation
All WPs	04/03/2016	1h30	NAIST, IIJ-II, NII, UT, IMT, ATOS, NASK, 6CURE	All WPs deliverables progress meeting; Video demonstrations; Next plenary preparation

6.5 Exchange of researchers and students

Within the framework of the *NECOMA* project, collaboration does not only occur through a common technical agenda, sharing data and collaborating on complementary tasks, but also through the exchange of skills and knowledge, and in particular by exchange of personnel. In this perspective, we sought to select adequate profiles, complementary to the hosting institution's skills, to jointly fulfil some project tasks.

6.5.1 Exchange of students

Exchanging students between partners of both consortia is a very effective method for the exploitation of knowledge and tools. Students are directly

tapped into *NECOMA* datasets, tools and algorithms to contribute to some specific tasks. Expected outputs, aside from task's completion, are knowledge transfer, international experience and academic publication. Indeed, students are highly motivated to prove their competence and acquire additional know-how and skills. Additionally, an international experience is often required in industry, while academia often assesses a student profile by his/her publication. Below, we report about the research stays of our students at another partner's institution.

Sirikarn PUKKAWANNA

Origin	NAIST
Host	IMT
Position	Ph.D student
Duration	3 months from Sep. to Dec. 2013
Task	T2.1.3: Endpoint threat data analysis
Subject	Classification of SSL servers
Results	1 publication at BADGERS 2014 [13] as well as several tools and recommendations.

Sirikarn contributed to the analysis of SSL servers responses dataset provided by IMT. She first attempted to analyze the data using techniques such as the S-transform, as developed during her thesis. When it resulted that the data were not appropriate for such techniques, she began delving into mining techniques with the help of an IMT Ph.D student, Dingqi YANG. A fruitful collaboration allowed them to develop several tools to query and analyze the SSL dataset. In particular, a query interface was developed over the Wombat API and three approaches were proposed to make sense of the SSL dataset:

1. a protocol version and encryption-algorithm-based classification: since SSL/TLS communication has been the victim of many attacks over the years, there are a number of protocol/encryption algorithm combinations which are highly advised against. Based on such knowledge, a score is computed, reflecting the vulnerability of connections established by the SSL server;

2. a Distinguished Name (DN)-based classification: since an SSL server certificate allows the authentication of a server, utmost attention should be given to the information displayed in the certificate. We consider that validated certificates were more trustworthy than self-signed certificates, classifying SSL servers as risky or seemingly harmless;

3. a combination of both approaches: a security assessment score is computed from a weighted sum of criteria from both assessments (vulner-

ability and trustworthiness) in order to define the level of security provided by the SSL server, ranging from worst to best.

Although the dataset comprised several collections of the whole IP range across time, we found no server providing the best security level, with half of them being actually classified as bad. Additionally, our study found no correlation between vulnerability and trustworthiness of the SSL servers, i.e., a server with a trusted certificate may provide an unsecure communication channel and vice versa. After surveying other related works, we proposed to extend the SSL server response collection to 45 features.

Vincent DANCHÉ

Origin	IMT
Host	NAIST and NII
Position	Master student
Duration	6 months from Apr. to Sep. 2014
Task	T2.1.2: Large-scale DNS traffic analysis and T2.1.4: Cross-layer threat data analysis
Subject	PCA analysis of DNS anomalies
Results	1 internship report and several tools.

Vincent has spent some time acquiring knowledge about machine learning and know-how about big data analytics in order to interact with *NECOMA*'s Hadoop cluster deployed by the Japanese consortium. His objective was to apply cross-layer analysis to datasets hosted at Japanese partners in order to better characterize DDoS attacks. With a large choice of datasets at hand, Vincent focused on DNS traffic data since DDoS reflection attacks often use DNS or other UDP-based protocols as their reflection medium. As DNS reflection attacks rely on spoofing queries from the victim host, Vincent studied traffic at 5 different DNS resolvers deployed at the University of Tokyo. He first tried to detect such anomalous events by computing an entropy score based on the distribution of queries among clients identified by their IP address, as given a suggestion to use by a NAIST Ph.D student, Sirikarn PUKKAWANNA. Vincent posited that when some suspicious users tend to spoof a small number of IP addresses and generate a lot of queries, the entropy score will approach zero. He then proposed a module to detect suspicious volumetric DNS query flows based on the entropy score computation. Finding out that his proposed module was vulnerable to some poisoning attack, he set out to further investigate the behavior of DNS traffic. During his stay at NII, he was suggested to use principal component analysis (PCA) to study the behavior of DNS traffic based on the features from the used dataset with the addition of the entropy score he previously proposed. His 7-variable PCA allowed the confirmation of the presence of anomalous events

in a sample dataset that actually contained suspected malicious events. By plotting the traffic captures along the principal component axes, Vincent could distinguish several clusters. He later used k-means clustering in order to automate the detection of anomalies within the different clusters, assuming that legitimate clusters that express general behavior usually vary close to the principal component axes. When representing the obtained cluster on the time-series representation of the DNS traffic capture, Vincent clearly visualized the anomalies he previously suspected. Manual analysis actually confirmed the presence of a DDoS event, as well as a scan and a spoofing attack.

Hirotaka FUJIWARA

Origin	NAIST
Host	IMT
Position	Master student
Duration	2 months from Sep. to Nov. 2014
Task	T2.1.3: Endpoint threat data analysis
Subject	Study of drive-by download attacks based on JS obfuscation features
Results	1 technical report [6] and several tools.

Hirotaka has been studying obfuscation features occurring in JavaScript code embedded in drive-by download attacks under the supervision of Gregory BLANC. His survey encompassed both legitimate sites, crawled from a list of popular websites ranked at Alexa, and malicious sites extracted from traffic captures made available during the MWS conference, a Japanese domestic symposium on malware research. He assumed that although obfuscation may not indicate malice, legitimate and malicious usage of obfuscation may differ. Based on several string features, he distinguished several types of obfuscation, as reported in existing literature [19], such as randomization, data obfuscation, encoding and logic structure obfuscation. In particular, he used the length of script lines as a major indicator: malicious scripts tend to include scripts in which code lines are under 600 characters. Malicious scripts will then usually deobfuscate its contents before issuing requests to some other domain controlled by the attacker, as it is usually the case in the context of a drive-by download attack. Hirotaka therefore posited that malicious domains must be obfuscated within the trap page code in order to evade signature-matching. Hence, he proposed a real-time drive-by download system where a proxy intercepts HTTP requests and responses, analyzing the presence of third-party domains in subsequent requests. If these third-party domains were not embedded in previous responses, it means they may have been concealed in the contents using obfuscation. Such behavior is assumed to be anomalous, if not malicious.

Wataru TSUDA

Origin	NAIST
Host	FORTH
Position	Master student
Duration	2 and half months from Oct. to Dec. 2014
Task	T2.1.2: Large-scale DNS traffic analysis and T2.1.4: Cross-layer threat data analysis
Subject	Detection of infected hosts through analysis of DNS responses
Results	1 technical report [17] and several tools

Wataru has been studying detection algorithms for diverse kinds of botnets through the development of feature vectors for analyzing DNS response packets and through the adaptation of multiple machine-learning algorithms that detects DGA-based botnets with high accuracy. His feature vectors incorporate information from multiple sources, including BGP prefixes from Team Cymru, IP-based reputation from DNSWL, malware domain blocklist, and the list of frequently visited sites as provided by Alexa. His study leveraged the analysis infrastructure and datasets offered by *NECOMA*.

His study has significantly benefited from the infrastructure and expertise of *NECOMA*, resulting in the improvement of detection accuracy, as well as the improved quality of information sources.

Daishi ITO

Origin	NAIST
Host	IMT
Position	Master student
Duration	4 months from Mar. to June 2015
Task	T3.3.1: DDoS mitigation based on software defined networks
Subject	Rule placement optimization in OpenFlow switches
Results	1 master's thesis

Daishi investigated potential technical issues that may arise during the application of SDN for DDoS mitigation purposes. Under the supervision of Dr. Gregory Blanc from IMT, he analyzed several papers for maintaining the integrity of OpenFlow networks under constant reconfiguration, as well as several papers on the performance implications of DDoS to the SDN. Inspired by existing papers on SDN-based distributed firewalls and access control, he continued to study access control issues in the OpenFlow networks, along with some possible optimization methods. After completing the exchange, he continued the study, resulting in his master's thesis on rule opti-

mization technique among OpenFlow network nodes, where he successfully demonstrated that his proposed technique can reduce 20% of unwanted traffic on a particular, hierarchical configuration of OpenFlow switches, using Ryu OpenFlow controller and his plug-in software module.

Pernelle MENSAH

Origin	IMT
Host	NAIST
Position	Master student
Duration	6 months from Apr. to Sep. 2015
Task	T1.4: Periodic test campaigns and T2.3.2: Development of threat analysis modules
Subject	Impact of SSL/TLS usage in HTTPS phishing incidents
Results	1 publication at BADGERS 2015 [10] as well as several tools

Pernelle originally joined NAIST to integrate tools developed by IMT into the threat analysis platform. During her stay at NAIST, Kazuya OKADA collaboratively supervised Pernelle, with the remote assistance of Gregory BLANC. She is the main implementer of the n6 interface that has made the SSL dataset hosted at IMT available to the public. During her internship, and even after completing it, she maintained the hosting server for several months. She has extended the work done by Sirikarn in a previous internship, in order to include her recommendations. This experience helped her analyze further HTTPS information, in particular with the ones collected at UT by Daisuke MIYAMOTO. This led to a joint work between IMT, NAIST and UT in order to characterize the excessive trust that users put in SSL/TLS artifacts, particularly in the case of phishing incidents. While a previous survey, carried out at NAIST by Sirikarn and another intern, proved that a specific class of HTTPS phishing websites were likely to bypass existing protection mechanisms, Pernelle offered a visual-hash-based solution to detect hacked domains hosting phishing pages.

6.5.2 Exchange of researchers

In addition to exchange of students, researchers are taking advantage of opportunities to engage in extended research interactions with several partner institutions. Long-term visit has been initially envisioned but turned out to be difficult due to other project engagements and budget constraints.

Thus far, the following exchange of researchers have taken place:

- Gregory Blanc, visiting Keio University in July 2013
- Youki Kadobayashi, visiting Telecom SudParis in September 2013

- Hajime Tazaki, visiting FORTH and Telecom SudParis in October 2014
- Youki Kadobayashi, visiting Telecom SudParis in October 2014
- Kazuya Okada, visiting Telecom SudParis in March 2015

7 Dissemination through demonstration videos

This chapter introduces a set of short videos that demonstrate the case studies and complements *NECOMA* Deliverable 4.2.

7.1 DDoS mitigation

7.1.1 Pushing defenses upstream

The demonstration video shows a scenario with simulated (replay of synthesized attack traffic) DNS amplification attack against a target over a small MPLS network. The attack saturates the network capacity and the mitigation mechanism is triggered. The mitigation mechanism uses attack sources IP addresses identified by local defenses at the target level and/or using external threat analysis data available via n6 API. The mechanism reconfigures upstream routers to send traffic from those sources to separate MPLS tunnels set up for suspicious traffic and with lower resource allocation than tunnels set up for legitimate traffic. This limits the amount of suspicious traffic reaching the target to a level that avoids congestion and can be handled locally.

7.1.2 SDN based DDoS mitigation

The demonstration video shows a simple SDN-based mitigation mechanism against Domain Name System (DNS) amplification attacks. The DNS-based amplification attack is a common type of DDoS attacks. The mechanism mitigates the attack on OpenFlow switches by dynamically reconfiguring the flow entries through the OpenFlow controller. Additionally, we also demonstrate a pipeline from data acquisition, threat detection, information sharing, and mitigation. Thus we consider that this video is fulfilling one of the important goals of *NECOMA*, the security pipeline.

7.1.3 DDoS mitigation as a service

The purpose of this demo is to illustrate how our proposed autonomic defense framework can systematically integrate several security functions together, ranging from traffic monitoring and detection to mitigation, in order to mitigate DDoS attacks without much human intervention. In particular, monitoring and anomaly detection run on the customer side, while the mitigation engine runs on the ISP side and is provided as an on-demand service to the customers. We choose video streaming service as protected assets and simulate various DDoS attacks, e.g., UDP flood, TCP flood, and ICMP flood, to test the feasibility and effectiveness in terms of a set of pre-defined QoE metrics. Technically, all traffic flows are tagged with FlowID (used to distinguish legitimate and malicious flows) and labels (associated with QoS and used to determine routing paths) at the ingress router of ISP. Traffic statistics are collected by an OF switch in the customer network, and a threshold-based anomaly detection scheme is employed to detect the anomalous traffic which has larger packet rate. Then the alerts, as soon as they are triggered, will be shared with ISP via SDN controller to controller communication. As a result, the ISP controller updates the label of those traffic flows which have the FlowID indicated in the alerts, eventually leading to their redirection to the middlebox.

7.2 Botnet introspection

7.2.1 C&C server introspection with DNS queries

This video shows the mechanism of our botnet introspection system using DNS Response Policy Zones (RPZ) and reverse web proxy. The Domain Generation Algorithm (DGA) employed by botnets uses dynamically generated domain name to connect to botnet C&C server. Our proposed system imitates IP addresses of these generated domain names to our data collection server. As a result, we could introspect C&C communication between malware and botnet. The demonstration also presents our malware traffic mitigation system which uses results of our analysis system.

7.3 Smartphone user protection

7.3.1 Drive by download prevention

This video demonstrates a browser-based prevention mechanism against drive-by-download attacks. The mechanism detects suspicious domain requests that are generated by obfuscated JavaScripts. Typically, obfuscation has been used to hide software internals from the others. JavaScript obfuscation is common due to the fact that JavaScript code must be down-

loaded to user's browser for execution. Recently, obfuscation has also been used to redirect user access towards malware downloads, in addition to its legitimate usage. The proposed mechanism detects undesired domain requests that are generated by obfuscated JavaScripts. The mechanism is implemented as a Chrome extension. When the Chrome extension detects suspicious domain accesses, visible alerts are raised to the user.

We have published the source code of the extension on our GitHub repository[1] with a user manual. You can freely try the extension on your laptop.

7.3.2 Phishing prevention

This video presents the personalization of a cyber defense system against phishing as the model threat, taking into account the users' awareness level. Within the context of personalization, we contend that there are different types of users: novices often fall victim to phishing, whereas experts can distinguish fraudulent websites from benign websites. We analyzed web users' eye movement, and subsequently developed a defense system for identifying the user type. The system, implemented as a browser's extension, selects the appropriate type of cyber defense: one is designed for experts, thus it is light-weight but has high precision. Another one is intended for novices, thus it comes with a deep-inspection and achieves higher recall. The video demonstrates the defense system, highlighting its performance in terms of precision, recall, mean time to delay, and usability, then describes its dissemination plans.

7.3.3 Smartphone firewall

This video introduces a firewall system for protecting smartphones from cyber threats. The key idea is to offload firewall functions to OpenFlow-capable wireless access points (APs). The widespread use of smartphones requires protection against cyber threats targeting such devices. *NECOMA* therefore explored the suitable protection methods, among which OpenFlow-capable APs are considered most appropriate, as they can facilitate configuring filtering rules and implement defense at the closest point to the device, while at the same time reducing energy consumption, thus avoiding heavy battery drain on smartphones due to firewall applications. In this video, we present our prototype implementation, then describe information pipelining in order to provide cyber defense based on threat information, and finally describe evaluation results.

[1] https://github.com/necoma/drive_by_download_protector

7.3.4 SMS fraud protection

This video introduces a way to protect mobile users from malicious applications that are trying to secretly send SMS messages (similar dial premium numbers), charge the users and produce large revenues for the malicious user that created that application. The key idea behind this protection mechanism is the following: When the malicious application is executed, it will try to send an SMS message. While trying to send this SMS message, our mechanism will display a dialog box to the user. The dialog informs the user of the message being sent, along with the destination number and offers three options. It is possible to either reject the message and blacklist the number, allow the message and whitelist the number or decide later. A CAPTCHA or a trivial math problem has to be solved to be able to pick any action.

7.4 Malware campaign mitigation

This video presents the detection, analysis and remediation process for a malware campaign. We show how the available datasets of malicious URLs can be analyzed to detect common patterns. For the most interesting patterns found we show how they can be analyzed to confirm that they are indeed malware campaigns and extract actionable information, such as C&C server addresses. Then we show how filtering rules (ready to apply in networks) can be generated and used. The video presents the full path from analysis of high-level threat information such as unfiltered URL collections to actual remediation actions.

8 Other dissemination activities

This chapter describes the other types of our dissemination activities.

8.1 Participation in exhibitions

8.1.1 Participation in FIC exhibition

NECOMA was presented at a stand on the 8th International Cybersecurity Forum[1], FIC, during January 26-27, 2016 in Lille, France. FIC is a security and privacy oriented event with strong presence from government agencies and industry.

During the event we probed the exploitation potential of our work in NECOMA in discussions with the exhibition participants including both current clients and future prospects. We also showcased our participation in NECOMA on our stand and used it as one of the arguments in hiring discussions as an example of the possibilities of engaging international collaboration and research we can offer.

8.1.2 Participation in INTEROP Tokyo 2015 exhibition

NECOMA presented SDN-based DDoS Defense in INTEROP Tokyo 2015, the leading global business technology event held in Makuhari Messe (Chiba) from June, 10-12, 2015, which is the largest event in Asia.

Through the exhibition, we demonstrated our cyberdefense system that can mitigate DDoS traffic flows. We also assessed the operability of the system by interconnecting many SDN-IX solutions. The audience was comprised of business people, engineers, and researchers.

In addition to that, we also demonstrated our prototype version in INTEROP Tokyo 2014, during June, 11-13, 2014.

[1] https://www.forum-fic.com/site/GB

8.2 Dissemination through educational activities

8.2.1 NECOMA summer school

During July, 29-30, 2015, *NECOMA* summer school was organized at Komaba Campus in the University of Tokyo. We intended to disseminate our tools and usage through students who applied to the class. We invited 10 students from diverse institutions within Japan: undergraduate school, graduate school, and technical college.

On the first day, we introduced our MATATABI system and conducted hands-on exercises of analyzing security big data. We also demonstrated the NECOMAtter systems and described how we utilize graph database for building an information sharing platform. On the second day, we let students implement cyber-defense programs that run on top of an OpenFlow controller. We then discussed the cognitive task analysis tools and its importance in cybersecurity by presenting EyeBit system.

8.2.2 Anti-phishing education

We developed an educational material for Japanese schoolchildren. This activity intended to study the importance of checking the address against phishing attacks. We performed a small study with students in three elementary schools in Japan from March 2015 to May 2015. It should be noted that the elementary school students do not own credit cards and do not have bank accounts. However, they have a lot of time to play online games, and their game accounts are traded in online black markets, known colloquially as real money trading.

8.2.3 FORTHcert training activities

FORTHcert is operated under the supervision of FORTH. FORTHcert is closely collaborating with other CERTs at the national and international levels. FORTHcert already has solutions in place for monitoring and detections mechanisms based on an Early Warning Intrusion System (EWIS). Extensive training has been performed by our team to the members of FORTHcert regarding the findings of the NECOMA project related to honeypots and how to improve the accuracy of their results. FORTHcert will improve the currently provided services to other organizations and SMEs (at the technical, operational and advisory levels).

9 Conclusion

The dissemination activities during the *NECOMA* project allowed us to reach the objectives, in accordance with the provisions of the dissemination plan defined at the beginning of the project. All major activities of the project were presented in detail. These include publications in journals, international conferences as well as two international workshops organized by *NECOMA*. In order to accelerate our dissemination activities, the key datasets used by *NECOMA* partners have been provided to stakeholders in academic and scientific communities. Most of our developed tools for constructing the framework, exchanging cyberthreat information, and for providing resilient cyberdefense have been published as open-source software. *NECOMA* has also participated in standards activities in IETF and ITU-T; our research results were also presented to stakeholdres in industrial communities.

The online dissemination activities were also carried forward to reach our international audiences. Our project websites, respectively organized by European and Japanese consortia, were regularly updated to disseminate the most recent news and provide information related to our research activities. The case studies were summarized in a set of videos and provided at the websites to highlight the achievements and insights from the *NECOMA* project. The websites will be updated as necessary to provide information on our relevant activities and research outcomes after completion of the project.

Precise international collaboration was needed to orchestrate the project team which is comprised of partners with different backgrounds. *NECOMA* therefore engaged in the exchange of knowledge through collaborative documentation work, which is facilitated by individual Work Package meeting. Each of the partners was encouraged to interact with novel technologies and concepts presented by the other partners. In order to induce the collaborative work, *NECOMA* partners have also exchanged researchers and students across the two consortia.

Bibliography

[1] The Simple Text Oriented Messaging Protocol. https://stomp.github.io/.

[2] K. Cho, K. Mitsuya, and A. Kato. Traffic data repository at the WIDE project. In *USENIX 2000 FREENIX Track*, San Diego, CA, June 2000.

[3] R. Danyliw, J. Meijer, and Y. Demchenko. RFC 5070: The incident object description exchange format, Dec. 2007.

[4] Electronic Frontier Foundation. https://www.eff.org/observatory.

[5] R. Fontugne, J. Mazel, and K. Fukuda. Hashdoop: A mapreduce framework for network anomaly detection. In *Computer Communications Workshops (INFOCOM WKSHPS), 2014 IEEE Conference on*, pages 494–499, April 2014.

[6] H. Fujiwara, G. Blanc, H. Hazeyama, and Y. Kadobayashi. A Study of Drive by Download Attack detection using the features of the obfuscation. Technical Report ICSS2014-60, IEICE, Nov. 2014.

[7] ITU-T FSTP-TACL. Telecommunications Accessibility Checklist, 2006.

[8] O. Levillain, A. Ébalard, B. Morin, and H. Debar. One year of ssl internet measurement. In *Proceedings of the 28th Annual Computer Security Applications Conference*, ACSAC '12, pages 11–20, New York, NY, USA, 2012. ACM.

[9] J. Lorchat, C. Pelsser, R. Bush, K. Shima, H. Schlesinger, and L. Johansson. TAMIAS: a distributed storage built on privacy and identity. In *The 28th Trans European Research and Education Networking Conference, 21 - 24 May, 2012, Reykjavik, Iceland*, May 2012.

[10] P. Mensah, G. Blanc, K. Okada, and Y. Kadobayashi. AJNA: Anti-Phishing JS-based Visual Analysis, to Mitigate Users' Excessive Trust in SSL/TLS. In *Proceedings of the 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*, Nov. 2015.

[11] D. Miyamoto, T. Iimura, G. Blanc, H. Tazaki, and Y. Kadobayashi. EyeBit: Eye-Tracking Approach for Enforcing Phishing Prevention Habits. In *Proceedings of the 3rd International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*, Sept. 2014.

[12] P. Pawliński, P. Jaroszewski, P. Kijewski, Ł. Siewierski, P. Jacewicz, and P. Zielony. Actionable information for security incident response. https://www.enisa.europa.eu/activities/cert/support/actionable-information, Nov 2014.

[13] S. Pukkawanna, Y. Kadobayashi, G. Blanc, J. Garcia-Alfaro, and H. Debar. Classification of SSL Servers based on their SSL Handshake for Automated Security Assessment. In *Proceedings of the 3rd International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*, Sept. 2014.

[14] R. Danyliw and J. Meijer and Y. Demchenko. RFC5070 - The Incident Object Description Exchange Format. http://www.ietf.org/rfc/rfc5070.txt.

[15] Y. Sekiya, T. Ishihara, and H. Tazaki. DNSSEC simulator for realistic estimation of deployment impacts. *IEICE Communications Express*, 3(10):305–310, 2014.

[16] H. Tazaki, K. Okada, Y. Sekiya, and Y. Kadobayashi. MATATABI: Multi-layer Threat Analysis Platform with Hadoop. In *3rd International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*, page 8, Sept. 2014.

[17] W. Tsuda, Y. Kadobayashi, H. Fujiwara, and S. Yamaguchi. Detecting infected hosts with machine learning analysis of DNS responses. Technical Report ICSS2014-62, IEICE, Nov. 2014.

[18] Z. Wilcox-O'Hearn and B. Warner. Tahoe: the Least-Authority FileSystem. In *Proceedings of the 4th ACM international workshop on Storage security and survivability*, pages 21–26. ACM, Oct. 2008.

[19] W. Xu, F. Zhang, and S. Zhu. The Power of Obfuscation Techniques in Malicious JavaScript Code: A Measurement Study. In *Proceedings of the 7th International Conference on Malicious and Unwanted Software (MALWARE)*, Oct. 2012.