TAL TECH

DOCTORAL THESIS

Automating Defences against Cyber Operations in Computer Networks

Mauno Pihelgas

TALLINNA TEHNIKAÜLIKOOL
TALLINN UNIVERSITY OF TECHNOLOGY
TALLINN 2021

TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
36/2021

Automating Defences against Cyber Operations in Computer Networks

MAUNO PIHELGAS

TAL TECH PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies
Department of Software Science

The dissertation was accepted for the defence of the degree of Doctor of Philosophy in Computer Science on 10 June 2021

Supervisor: Dr. Risto Vaarandi,
Department of Software Science, School of Information Technologies,
Tallinn University of Technology,
Tallinn, Estonia

Co-supervisor: Professor Dr. Olaf Manuel Maennel,
Department of Software Science, School of Information Technologies,
Tallinn University of Technology,
Tallinn, Estonia

Opponents: Professor Dr. Anja Feldmann,
Max Planck Institute for Informatics,
Saarbrücken, Germany

Professor Dr. Jan Vykopal,
Masaryk University,
Brno, Czechia

Defence of the thesis: 11 August 2021, Tallinn

Declaration:
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology, has not been submitted for any academic degree elsewhere.

Mauno Pihelgas _____
signature

ISSN 2585-6898 (publication)
ISBN 978-9949-83-718-2 (publication)
ISSN 2585-6901 (PDF)
ISBN 978-9949-83-719-9 (PDF)
Printed by Koopia Niini & Rauam

Full text https://digikogu.taltech.ee/et/Item/beb3e841-9c6e-4496-a73a-17148bc941ef

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
36/2021

Arvutivõrkude kaitse automatiseerimine küberoperatsioonide vastu

MAUNO PIHELGAS

TAL TECH
KIRJASTUS

Contents

List of Publications

I R. Vaarandi and M. Pihelgas. Using Security Logs for Collecting and Reporting Technical Security Metrics. In *Military Communications Conference (MILCOM), 2014 IEEE*, pages 294–299, October 2014

II R. Vaarandi and M. Pihelgas. LogCluster - A data clustering and pattern mining algorithm for event logs. In *Network and Service Management (CNSM), 2015 11th International Conference on*, pages 1–7, November 2015

III R. Vaarandi, M. Kont, and M. Pihelgas. Event log analysis with the LogCluster tool. In *MILCOM 2016 - 2016 IEEE Military Communications Conference*, pages 982–987, November 2016

IV B. Blumbergs, M. Pihelgas, M. Kont, O. Maennel, and R. Vaarandi. Creating and Detecting IPv6 Transition Mechanism-Based Information Exfiltration Covert Channels. In *Secure IT Systems: 21st Nordic Conference, NordSec 2016, Oulu, Finland, November 2-4, 2016. Proceedings*, pages 85–100. Springer International Publishing, 2016

V M. Kont, M. Pihelgas, K. Maennel, B. Blumbergs, and T. Lepik. Frankenstack: Toward real-time Red Team feedback. In *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*, pages 400–405, October 2017

VI M. Pihelgas. Design and Implementation of an Availability Scoring System for Cyber Defence Exercises. In *14th International Conference on Cyber Warfare and Security (ICCWS 2019)*, page 329–337, 2019

VII P. Théron, A. Kott, M. Drašar, K. Rzadca, B. LeBlanc, M. Pihelgas, L. Mancini, and A. Panico. Towards an active, autonomous and intelligent cyber defense of military systems: The NATO AICA reference architecture. In *2018 International Conference on Military Communications and Information Systems (ICMCIS)*, pages 1–9, May 2018

VIII P. Théron, A. Kott, M. Drašar, K. Rzadca, B. LeBlanc, M. Pihelgas, L. Mancini, and F. de Gaspari. Reference Architecture of an Autonomous Agent for Cyber Defense of Complex Military Systems. In *Adaptive Autonomous Secure Cyber Systems*, pages 1–21, Cham, 2020. Springer International Publishing

IX A. Kott, P. Théron, L. V. Mancini, E. Dushku, A. Panico, M. Drašar, B. LeBlanc, P. Losiewicz, A. Guarino, M. Pihelgas, and K. Rzadca. An introductory preview of Autonomous Intelligent Cyber-defense Agent reference architecture, release 2.0. *The Journal of Defense Modeling and Simulation*, 17(1):51–54, 2020

X R. Vaarandi and M. Pihelgas. NetFlow Based Framework for Identifying Anomalous End User Nodes. In *15th International Conference on Cyber Warfare and Security (ICCWS 2020)*, page 448–456, 2020

XI M. Pihelgas and M. Kont. Frankenstack: Real-time Cyberattack Detection and Feedback System for Technical Cyber Exercises. In *2021 IEEE CSR Workshop on Cyber Ranges and Security Training (CRST)*. IEEE, July 2021. (Accepted paper)

Author's Contributions to the Publications

I I conducted the background study on security metrics research, properties of good security metrics, and how to create an organisational security metrics program. Furthermore, I offered proposals and suggestions in the metrics selection process as well as helped draw conclusions on the example security metrics extracted from the organisational framework implementation.

II I conducted performance evaluation experiments of the LogCluster algorithm on Locked Shields data including both the final and intermediate versions of the algorithm. I also participated in the design and development discussions of the LogCluster algorithm. During the development phase of the algorithm, I conducted a detailed evaluation of various data structures used by the cluster candidate support aggregation procedure, measuring how different data structures influence the algorithm performance under heavy workloads.

III I participated in the discussions and helped create several novel LogCluster usage examples presented in the paper.

IV I was responsible for describing the data exfiltration detection experiment and assembling the set of capture combinations of each exfiltration tool, source and destination port number, transport layer protocol, and IP version. Altogether, 126 unique packet capture (PCAP) files were generated for analysis. We used dedicated monitoring nodes to run a selection of five popular open-source detection tools which would analyse each of these PCAP files, produce flow records, and potentially generate alerts for suspicious activity. We presented the detection results in an extensive table and I provided the detailed discussion of the detection results.

V I was the co-author of the cyberattack detection and feedback system designed to increase the training benefit for the cyber red team participants during cyber exercises. The authors organised a one-week hackathon during the Crossed Swords 2017 Test Run for sprint-developing the Frankenstack framework—the near real-time technical feedback system for the cyber red team. I was responsible for designing the post-processing, filtering and correlation engine to process raw events from intrusion detection system, syslog, Windows logs. These transformed and correlated (i.e., meaningful) events were then displayed on feedback dashboard for the cyber red team to see and learn. The framework was successfully implemented and tested in the NATO CCD COE's technical cyber security exercise Crossed Swords. Furthermore, I provided input for the cyber red team situation awareness (SA) feedback questionnaire prior to the event and later analysed the feedback to extract valuable takeaways for improving Frankenstack in the upcoming years.

VI I was the sole author of this paper and have been the primary designer, developer and implementer of the Availability Scoring system for the Locked Shields cyber defence exercise since 2014. The paper describes the design process and reasoning behind various decisions made in the scoring system that has been used annually since 2014. The system provides automated scoring checks for a variety of systems throughout the exercise. While the framework employs many standard system monitoring practices, one of the primary differentiators is the adversarial environment in which the service availability checks have to be performed. It is in the winning interest of participating teams to mislead the monitoring system in order to obtain a better score for the availability of services.

VII This is the first of three publications on the topic of **A**utonomous **I**ntelligent **C**yber-defence **A**gents (AICA). Publication VII is based on the intermediate results of the work done by the NATO Science and Technology Organisation's IST-152 Research Task Group (RTG) between 2016 and 2018. I was an active member of this research group. Publication VII captures the primary concepts of the initial reference architecture [85] published by the IST-152 group. This preliminary research resulted in the publication of the *Initial Reference Architecture of an Intelligent Autonomous Agent for Cyber Defense* [85], in which I directly contributed to the research and write-up of the *Sensing and World State Identification* chapter. Correspondingly, I was the co-author of the same chapter within Publication VII.

VIII This book chapter published in [68] is the second of three publications on the topic of AICAs. This chapter is largely based on the second revision of the *Autonomous Intelligent Cyber-defense Agent (AICA) Reference Architecture* report [86] that comprises the work done by the NATO IST-152 group between 2016 and 2019. In Publication VIII, I contributed to the research and write-up of section *Sensing and World State Identification* and was the primary author of section *Use Cases* that provides an example scenario of AICAs being deployed within Unmanned Aerial Vehicles (UAV).

IX This journal article is the third of three publications on the topic of AICAs. Publication IX serves as an introductory and overview article of the second revision of the *Autonomous Intelligent Cyber-defense Agent (AICA) Reference Architecture* report [86] that concludes the work done by the NATO IST-152 RTG between 2016 and 2019. Within this second release of the reference architecture, I directly contributed to the research and write-up of the *Rationale of AICA and Scenario* and *Sensing and World State Identification* chapters.

X I provided input for the development and features of the novel detection framework. In later phases, I supported the work by looking through the analysed NetFlow dataset to provide additional observations on the performance of the proposed methods.

XI I was the primary author of this paper. The paper describes the research and development effort of Frankenstack following the initial paper (Publication V) in 2017. This latest publication describes the updated architecture, event normalisation, data enrichment methods, and improved event processing within the newly developed version. In addition to contributing most of the manuscript write-up, I was responsible for the development of the *Exercise asset collection* and *Python event shipper* modules described in the paper.

Abstract

Automating Defences against Cyber Operations in Computer Networks

This thesis is based on a collection of eleven publications. The thesis explores the improvement of organisational security monitoring capability and readiness to advance towards intelligent autonomous cyber defence systems.

Additionally, the thesis aims to reduce the gap between suggestions derived from academic research and practical guidelines that are useful for cyber defenders. The feasibility of utilising theoretical research outcomes in practice has been criticised in related publications by several different authors. To relieve this issue, this thesis and the bundled collection of publications provide numerous actionable recommendations and practical examples.

This thesis addresses problems in the areas of establishing which metrics are relevant for security monitoring, how to build both general-purpose and cyber-exercise-specific situation awareness systems, how to raise SA qualifications and readiness of cyber defenders, how to implement and verify novel log mining algorithms and network security frameworks for cyber defence, and how to improve cyber defences by designing autonomous intelligent cyber-defence agents.

The thesis provides recommendations for metrics and log data collection, transformation, and analysis methods alongside relevant data representation techniques. Furthermore, a novel data clustering and log mining algorithm LogCluster is proposed, compared thoroughly with several other log analysis tools, and later used to provide practical examples of clustering logs from two different cyber security exercises (Locked Shields and Crossed Swords). Furthermore, the thesis describes two novel cyber-exercise-specific situation awareness systems—Frankenstack and the Availability Scoring system: comprising an overview of the development process, technical architecture, and validation during the aforementioned cyber security exercises. In the area of network security, the thesis describes the research on data exfiltration detection with open-source tools and details a novel NetFlow-based anomaly detection framework. Finally, the concept and reference architecture for autonomous intelligent cyber-defence agents is described and proposed as the basis for future military and civil cyber defence systems.

Kokkuvõte
Arvutivõrkude kaitse automatiseerimine küberoperatsioonide vastu

Käesolev ingliskeelne doktoritöö põhineb autori üheteistkümnel publikatsioonil ja uurib võimalusi arvutivõrkude küberkaitse automatiseerimiseks küberoperatsioonide vastu. Töö peamine eesmärk on luua eeldused ning tõsta üldist valmisolekut autonoomsete küberkaitse süsteemide arendamiseks ja juurutamiseks lähitulevikus.

Töö teine eesmärk on teoreetilise teadustöö ning praktikas rakendatavate juhiste tihedam sidumine. Akadeemilistes publikatsioonides jagatud soovituste rakendamine on tihti liiga keeruline—taoliste soovituste ebapraktilisust on kritiseeritud mitmes doktoritöös viidatud allikas. Selle probleemi leevendamiseks pakub käesolev doktoritöö ja sellega kaasnevad publikatsioonid arvukalt praktilisi soovitusi ning näiteid tehniliste lahenduste juurutamiseks.

Doktoritöös otsitakse vastuseid järgnevatele küsimustele ja probleemidele: milliseid tehnilisi meetrikaid on oluline jälgida küberturbe seires; kuidas rajada nii tavakasutuse kui ka küberharjutuste jaoks mõeldud situatsiooniteadlikkuse süsteeme; mil viisil oleks võimalik tõsta küberkaitsjate üldist kvalifikatsiooni ja treenida nende oskusi situatsiooniteadlikkuse valdkonnas; kuidas oleks võimalik testida uudsete andmekaevandamisalgoritmide ja võrguturbesüsteemide efektiivsust ja töökindlust; ning kuidas tõsta küberkaitse võimekust iseõppivate autonoomsete küberkaitse agentidega.

Doktoritöö kätkeb soovitusi meetrikate ja logiandmete kogumise, töötlemise ning esitlemise parendamiseks. Töö kirjeldab logide kaevandamise algoritmi LogCluster, võrdleb LogClusterit mitme konkureeriva logianalüüsi tööriistaga ning toob mitmeid praktilisi näiteid LogClusteri kasutamisest küberharjutuste andmekogude analüüsimiseks. Eraldi käsitletakse küberharjutuste tarbeks loodud kahte vabatarkvaralist monitooringusüsteemi: töö sisaldab ülevaadet nende süsteemide väljatöötamisest, komponentide tehnilisest ülesehitusest ning katsetamisest kahe erineva küberharjutuse, Locked Shields ja Crossed Swords, raames. Võrguturbe valdkonnas uurib töö andmelekete avastamist vabatarkvaraliste vahenditega ning kirjeldab hiljuti publitseeritud NetFlow-põhist võrguanomaaliate tuvastamise seireraamistikku. Viimaks kirjeldatakse kontseptuaalset intelligentsete küberagentide etalonarhitektuuri, mida saaks potentsiaalselt rakendada autonoomsete küberkaitse agentide arendamisel ja juurutamisel.

Curriculum Vitae

1. Personal data

Name	Mauno Pihelgas
Date and place of birth	3 January 1988, Haapsalu, Estonia
Nationality	Estonian

2. Contact information

Address	Tallinn University of Technology, School of Information Technologies, Department of Software Science, Ehitajate tee 5, 19086 Tallinn, Estonia
E-mail	info[at]pihelgas.eu

3. Education

2014–2021	Tallinn University of Technology, School of Information Technologies, Cyber Security, PhD studies
2010–2012	Tallinn University of Technology, Faculty of Information Technology, Cyber Security, MSc *cum laude*
2010–2012	University of Tartu, Faculty of Science and Technology, Cyber Security, MSc *cum laude*
2007–2010	Estonian Information Technology College, IT Systems Development, Diploma *cum laude*

4. Language competence

Estonian	native
English	fluent
Russian	basic level
German	basic level

5. Professional employment

2013– …	NATO Cooperative Cyber Defence Centre of Excellence, Technology Researcher
2012–2020	Tallinn University of Technology, Computer Lab Assistant
2010–2013	Elion Ettevõtted AS, Monitoring Administrator
2008–2010	Microlink Eesti AS, Data Centre Duty Technician

6. Certifications

2020–2023	Red Hat Certified Specialist in Advanced Automation: Ansible Best Practices
2017–2023	Red Hat Certificate of Expertise in Ansible Automation Exam (Ansible 2)
2016–2024	GIAC Continuous Monitoring Certification (GMON)
2014–2023	Red Hat Certified System Administrator (RHEL7)
2014–2023	Red Hat Certified Engineer (RHEL7)

7. Voluntary work

2011	Vaata Maailma Foundation, Restoration of donated computers for charity
2012–2020	Robotex, Sumo Robot workshop instructor

8. Computer skills

- Operating systems: GNU/Linux, MS Windows
- Document preparation: Vim, MS Code, LaTeX, Libre Office, MS Word
- Programming languages: Python, Bash, PHP, Perl, Go
- Scientific packages: Jupyter Notebooks, JupyterLab

9. Defended theses

- 2012, A Comparative Analysis of Open-Source Intrusion Detection Systems, MSc, supervisor Dr. Risto Vaarandi, Tallinn University of Technology
- 2010, Expanding Functionality of the Robot Control Platform of The Estonian Information Technology College, Diploma, supervisor Margus Ernits, Estonian Information Technology College

10. Field of research

- Security Monitoring
- Situation Awareness
- Cyber Security Exercises
- Log Analysis

11. Scientific work
Papers

1. R. Vaarandi and M. Pihelgas. Using Security Logs for Collecting and Reporting Technical Security Metrics. In *Military Communications Conference (MILCOM), 2014 IEEE*, pages 294–299, October 2014
2. R. Vaarandi and M. Pihelgas. LogCluster - A data clustering and pattern mining algorithm for event logs. In *Network and Service Management (CNSM), 2015 11th International Conference on*, pages 1–7, November 2015
3. R. Vaarandi, M. Kont, and M. Pihelgas. Event log analysis with the LogCluster tool. In *MILCOM 2016 - 2016 IEEE Military Communications Conference*, pages 982–987, November 2016
4. B. Blumbergs, M. Pihelgas, M. Kont, O. Maennel, and R. Vaarandi. Creating and Detecting IPv6 Transition Mechanism-Based Information Exfiltration Covert Channels. In *Secure IT Systems: 21st Nordic Conference, NordSec 2016, Oulu, Finland, November 2-4, 2016. Proceedings*, pages 85–100. Springer International Publishing, 2016
5. M. Kont, M. Pihelgas, K. Maennel, B. Blumbergs, and T. Lepik. Frankenstack: Toward real-time Red Team feedback. In *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*, pages 400–405, October 2017
6. M. Pihelgas. Design and Implementation of an Availability Scoring System for Cyber Defence Exercises. In *14th International Conference on Cyber Warfare and Security (ICCWS 2019)*, page 329–337, 2019

7. P. Théron, A. Kott, M. Drašar, K. Rzadca, B. LeBlanc, M. Pihelgas, L. Mancini, and A. Panico. Towards an active, autonomous and intelligent cyber defense of military systems: The NATO AICA reference architecture. In *2018 International Conference on Military Communications and Information Systems (ICMCIS)*, pages 1–9, May 2018

8. P. Théron, A. Kott, M. Drašar, K. Rzadca, B. LeBlanc, M. Pihelgas, L. Mancini, and F. de Gaspari. Reference Architecture of an Autonomous Agent for Cyber Defense of Complex Military Systems. In *Adaptive Autonomous Secure Cyber Systems*, pages 1–21, Cham, 2020. Springer International Publishing

9. A. Kott, P. Théron, L. V. Mancini, E. Dushku, A. Panico, M. Drašar, B. LeBlanc, P. Losiewicz, A. Guarino, M. Pihelgas, and K. Rzadca. An introductory preview of Autonomous Intelligent Cyber-defense Agent reference architecture, release 2.0. *The Journal of Defense Modeling and Simulation*, 17(1):51–54, 2020

10. R. Vaarandi and M. Pihelgas. NetFlow Based Framework for Identifying Anomalous End User Nodes. In *15th International Conference on Cyber Warfare and Security (ICCWS 2020)*, page 448–456, 2020

11. M. Pihelgas and M. Kont. Frankenstack: Real-time Cyberattack Detection and Feedback System for Technical Cyber Exercises. In *2021 IEEE CSR Workshop on Cyber Ranges and Security Training (CRST)*. IEEE, July 2021. (Accepted paper)

Conference presentations

1. **M. Pihelgas**. 'Security Metrics - Background Study and Suggestions for Organizational Networks and IT Systems', SAS-106 Symposium on Analysis Support to Decision Making in Cyber Defence & Security: 9–10 June 2014, Tallinn, Estonia

2. B. Blumbergs, **M. Pihelgas**. 'Creating and Detecting IPv6 Transition Mechanism-Based Information Exfiltration Covert Channels', 21st Nordic Conference on Secure IT Systems (Nordsec 2016): 2–4 November 2016, Oulu, Finland

3. **M. Pihelgas**, M. Kont. 'Hedghehog in the Fog: Creating and Detecting IPv6 Transition Mechanism-Based Information Exfiltration Covert Channels', CERT-EE Symposium 2017: 29–30 May 2017, Tallinn, Estonia

4. **M. Pihelgas**. 'Scoring a Technical Cyber Defense Exercise with Nagios and Selenium', 13th Open Source Monitoring Conference (OSMC 2018): 5–8 November 2018, Nuremberg, Germany

5. M. Kont, **M. Pihelgas**. 'Frankenstack. Busting the Red Team with Duct Tape, Spit and Tears', 5th Annual SuriCon: 30 October–01 November 2019, Amsterdam, The Netherlands

6. **M. Pihelgas**. 'Frankenstack: Real-time Cyberattack Detection and Feedback System for Technical Cyber Exercises', 2021 IEEE CSR Workshop on Cyber Ranges and Security Training (CRST): 26 July 2021, Online (Accepted presentation)

Elulookirjeldus

1. Isikuandmed

Nimi	Mauno Pihelgas
Sünniaeg ja -koht	03.01.1988, Haapsalu, Eesti
Kodakondsus	Eesti

2. Kontaktandmed

Aadress	Tallinna Tehnikaülikool, Tarkvarateaduse Instituut, Ehitajate tee 5, 19086 Tallinn, Estonia
E-post	info[ät]pihelgas.eu

3. Haridus

2014–2021	Tallinna Tehnikaülikool, infotehnoloogia teaduskond, Küberkaitse, doktoriõpe
2010–2012	Tallinna Tehnikaülikool, infotehnoloogia teaduskond, Küberkaitse, MSc *cum laude*
2010–2012	Tartu ülikool, Matemaatika-informaatika teaduskond, Küberkaitse, MSc *cum laude*
2007–2010	Eesti Infotehnoloogia Kolledž, IT süsteemide arendus, rakenduskõrgharidus

4. Keelteoskus

eesti keel	emakeel
inglise keel	kõrgtase
vene keel	algtase
saksa keel	algtase

5. Teenistuskäik

2013– …	NATO Küberkaitse Koostöö Kompetentsikeskus, Teadur-nõunik
2012–2020	Tallinna Tehnikaülikool, Praktikumi assistent
2010–2013	Elion Ettevõtted AS, Monitooringu administraator
2008–2010	Microlink Eesti AS, Serverikeskuste valvetehnik

6. Erialased sertifikaadid

2020–2023	Red Hat Certified Specialist in Advanced Automation: Ansible Best Practices
2017–2021	Red Hat Certificate of Expertise in Ansible Automation Exam (Ansible 2)
2016–2024	GIAC Continuous Monitoring Certification (GMON)
2014–2023	Red Hat Certified System Administrator (RHEL7)
2014–2023	Red Hat Certified Engineer (RHEL7)

7. Vabatahtlik töö

2011	Vaata Maailma Foundation, Vanade arvutite taastamine heategevuseks
2012–2020	Robotex, Sumorobotite töötoa juhendaja

8. Computer skills

- Operatsioonisüsteemid: GNU/Linux, MS Windows
- Kontoritarkvara: Vim, MS Code, LaTeX, Libre Office, MS Word
- Programmeerimiskeeled: Python, Bash, PHP, Perl, Go
- Teadustarkvara paketid: Jupyter Notebooks, JupyterLab

9. Kaitstud lõputööd

- 2012, Võrdlusanalüüs vabatarkvaralistest ründetuvastussüsteemidest, MSc, juhendaja Dr. Risto Vaarandi, Tallinna Tehnikaülikool
- 2010, Eesti Infotehnoloogia Kolledži roboti juhtimisplatvormi funktsionaalsuse laiendamine, juhendaja Margus Ernits, Eesti Infotehnoloogia Kolledž

10. Teadustöö põhisuunad

- Monitooring
- Situatsiooniteadlikkus
- Küberharjutused
- Logianalüüs

11. Teadustegevus
Teadusartiklite, konverentsiteeside ja konverentsiettekannete loetelu on toodud inglise-keelse elulookirjelduse juures.

ISSN 2585-6901 (PDF)
ISBN 978-9949-83-719-9 (PDF)