TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Karl Rikkonen 232666IVCM

From Noise to Knowledge: Developing a Machine Learning-Based Method for Threat Actor Infrastructure Attribution

Master's thesis

Supervisor: Hayretdin Bahşi
PhD
Co-supervisor Marvin Uku
MSc

Tallinn 2025

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Karl Rikkonen 232666IVCM

Masinõppepõhise meetodi väljatöötamine ohusubjekti IT-taristu omistamiseks

Magistritöö

Juhendaja: Hayretdin Bahşi
PhD
Kaasjuhendaja Marvin Uku
MSc

Tallinn 2025

Abstract

Attributing threat actor infrastructure in cybersecurity is essential yet challenging due to the complexity and volume of data available from the internet. It is a strategically critical task, as accurate attribution supports informed political decision-making at both national and organizational levels. Traditional methods, which depend on threat intelligence analysts examining indicators of compromise, often lack scalability and can miss subtle yet significant patterns linked to more advanced threat actors. This thesis addresses these challenges by proposing a machine learning-based method to turn complex and noisy data into actionable insights and knowledge for attributing threat actor infrastructure.

The developed methodology utilizes unsupervised clustering techniques to identify meaningful patterns that reflect distinct characteristics of infrastructure associated with threat actors. Applying quantitative evaluations and interpretability analyses helps security analysts derive clear, actionable intelligence from extensive datasets. The method is evaluated through experimentation, demonstrating its potential to enhance cybersecurity attribution capabilities substantially. The present thesis highlights the practical benefits and potential applications of machine learning in cybersecurity, emphasizing its role in improving threat actor attribution's speed, accuracy, and consistency.

This thesis is written in English and is 65 pages long, including 6 chapters, 9 figures, and 5 tables.

Annotatsioon

Masinõppepõhise meetodi väljatöötamine ohusubjekti IT-taristu omistamiseks

Küberkurjategijate IT-taristu omistamine kindlale isikule või organisatsioonile on küberjulgeolekus hädavajalik, et kohandada poliitilist hoiakut nii riiklikul kui ka ettevõtte tasandil. Omistamist aga raskendab internetist kogutava andmestiku suur maht ja ebaühtlus. Traditsioonilised analüüsipõhised võtted, mis tuginevad turvarikkemärkide (IP-aadressid, sertifikaatide räsid, jpt) käsitsi läbivaatusele, ei ole kasvuvõimelised ning võivad jätta märkamata lineaarsed seosed kinnisründeohtude ja küberkurjategijate vahel.

Käesolev magistritöö pakub lahenduseks masinõppel põhineva meetodi, mis teisendab keeruka ja mürarikka andmestiku kiiresti kasutatavaks teadmuseks, küberkurjategijate IT-taristu omistamiseks. Väljatöötatud metoodika rakendab juhenduseta masinõppe meetodeid (klasteranalüüsi), et tuvastada mustreid andmestikest, mis peegeldavad ründaja-spetsiifilisi tunnuseid. Kvantitatiivne hindamine ja klastrite seletatav analüüs võimaldavad turbeanalüütikutel eraldada ulatuslikest andmehulkadest selge ja usaldusväärse küberluureinfo. Eksperimendi tulemused näitavad, et meetod suurendab omistamise kiirust, täpsust ja järjepidavust, tõstes seeläbi organisatsioonide suutlikkust siduda IT-taristu objekte konkreetsete küberründajatega või nende poolt kasutatavate tööriistadega. Käesolev töö osutab masinõppe praktilisele väärtusele küberjulgeolekus ning selle võimele parandada küberkurjategija IT-taristu tõhusat omistamist.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 65 leheküljel, 6 peatükki, 9 joonist, 5 tabelit.