2019 /2500M-0101

ISSN 1797-3457 (verkkojulkaisu)
ISBN (PDF) 978-951-663-078-9

TIIVISTELMÄRAPORTTI

Tekoälyn käyttö poikkeamapohjaiseen tunkeutumisten havainnointiin verkkoliikenteestä

Tero Kokkonen, Jyväskylän ammattikorkeakoulu IT-instituutti, tero.kokkonen@jamk.fi
Samir Puuska, Jyväskylän ammattikorkeakoulu IT-instituutti, samir.puuska@jamk.fi
Janne Alatalo, Jyväskylän ammattikorkeakoulu IT-instituutti, janne.alatalo@jamk.fi
Eppu Heilimo, Jyväskylän ammattikorkeakoulu IT-instituutti, eppu.heilimo@jamk.fi
Antti Mäkelä, Jyväskylän ammattikorkeakoulu IT-instituutti, antti.makela@jamk.fi

Tiivistelmä. Verkotettujen tietojärjestelmien määrän kasvun, sekä kiihtyvän digitalisaation myötä yhteiskuntamme on entistä riippuvaisempi tietoverkoista ja niiden turvallisuudesta. Tämä koskettaa Puolustusvoimien lisäksi kaikkia yhteiskunnan turvallisuusviranomaisia, joiden johtamisjärjestelmät ovat riippuvaisia toimivista tietoverkoista ja tietojärjestelmistä. Puolustusvoimien tapauksessa tällaisia järjestelmiä ovat esimerkiksi tulenkäytön johtamisjärjestelmät. Erilaisten verkko- ja kyberhyökkäysten, tai yleisemmin tunkeutumisten määrä kriittisiä järjestelmiä kohtaan kasvaa jatkuvasti. Tämä kasvu luo tarpeen kokonaisvaltaiselle tunkeutumisten havainnoinnin kehitykselle ja tutkimukselle.
Kaikessa turvallisuustoimintaan liittyvässä päätöksenteossa tilannekuvalla ja tilannekuvan perusteella saavutettavalla tilannetietoisuudella on erittäin suuri arvo. Kyberturvallisuudessa tietoverkot ja verkotetut järjestelmät eivät noudata perinteisiä fysikaalisen maailman rajoituksia ja lainalaisuuksia, joten tilannekuvalla ja havaintojen visualisoinnilla on kyberturvallisuudessa entistäkin suuremmat vaatimukset ja haasteet.
Tutkimushankkeen lopputuloksena kehitettiin tekoälypohjainen tunkeutumisten havainnointijärjestelmä. Järjestelmän toiminta testattiin käyttämällä Kansallisen Kyberharjoituksen (KYHA18) aikana luotua tietoliikennekaappausta. Lisäksi käytettiin vapaasti saatavilla olevaa testidata-aineistoa, jonka perusteella taataan vertailtavuus kansainvälisiin tutkimustuloksiin. Tutkimuksessa kehitettiin lisäksi sovellus havaintojen ja toiminnan visualisointiin päätöksenteon tueksi. Tulosten pohjalta on tehty kaksi kansainvälistä tieteellistä julkaisua.

1. Johdanto

Nykyisin lähes kaikki tietojärjestelmät ovat verkotettuja, joten niihin kohdistuvien hyökkäysten sekä tunkeutumisten määrä kasvaa jatkuvasti. Tämän vuoksi tarvitaan kyvykkyys tunnistaa tunkeutumisia tietoverkoista ja järjestelmistä. Tekoa tai menetelmää, joka vaarantaa tietojärjestelmän tai –verkon toiminnan, kutsutaan yleisesti termillä ”tunkeutuminen” (intrusion). Tunkeutumisten havainnointijärjestelmistä käytetään yleisesti kansainvälistä nimitystä Intrusion Detection System (IDS).

Nykyaikaisessa kybertoimintaympäristössä tunkeutumisten osalta voidaan tunnistaa selkeä tarve monipuolisen sensorikyvyn kehittämiselle. IDS-järjestelmät jaetaankin kahteen kategoriaa niiden sijainnin mukaan: laitekohtaisiin (Host Intrusion Detection System, HIDS) ja verkkokohtaisiin (Network Intrusion Detection System, NIDS). Toiminnan mukaan IDS –järjestelmät jaetaan pääsääntöisesti kahteen eri kategoriaan; väärinkäytösten havainnointiin perustuviin havainnointijärjestelmiin (Misuse Based IDS) ja poikkeamien havainnointiin perustuviin havainnointijärjestelmiin (Anomaly Based IDS).

Tässä tutkimuksessa kehitettiin poikkeamien havainnointiin perustuva verkkokohtainen IDS-järjestelmä. Poikkeamien havainnointiin perustuville järjestelmille opetetaan verkkoliikenteen normaalimalli, jonka pohjalta järjestelmät pyrkivät havainnoimaan

poikkeamia tästä normaalimallista. Hyvänä puolena tässä on se, että niillä kyetään löytämään ennalta tuntemattomia hyökkäyksiä ns. nollapäivähyökkäyksiä, mutta heikkoutena on, että yleensä ne generoivat suuren määrän vääriä hälytyksiä (False Positive).

Tässä tutkimuksessa verkkoliikenteen normaalimallin oppimiseen ja toisaalta poikkeamien havainnointiin normaalimallista käytettiin syväoppivaa tekoälyä. Syväoppimiseen pohjautuva tekoälymalli ja koko IDS-järjestelmä kehitettiin käyttäen OpenSource ohjelmistokomponentteja.

2. Tutkimuksen tavoite ja suunnitelma

Vuoden 2018 aikana Jyväskylän ammattikorkeakoulun IT-instituutti on tehnyt samalla otsikolla MATINE tutkimusta, joka tuolloin suunniteltiin kaksivuotiseksi. Vuoden 2018 aikana tehty tutkimus on osoittanut erittäin lupaavia tuloksia tekoälyn hyödyntämiselle kyseisen ongelman ratkaisussa. Erityisesti projektin alussa mallinnettu organisaatioympäristö mahdollisti realistisen verkko- ja sovellusympäristön käytön, mukaan lukien aitojen haittaohjelmien ja hyökkäysvektorien käytön tutkimuksessa. Tässä jatkotutkimuksessa tekoälyyn ja koneoppimiseen perustuvan tunkeutumisten-havainnointijärjestelmäsovelluksen suorituskykyä parannettiin uudella koneoppimisalgoritmilla. Tämän lisäksi kehitettiin tulosten visualisointia parantamaan tilannekuvaa havaituista tunkeutumisista sekä kehitetyn mallin toimivuudesta.

3. Aineisto ja menetelmät

Tutkimusmetodologiana käytettiin konstruktiivista tutkimusta. Konstruktiivisen tutkimuksen lähtökohtana on rakentaa konstruktio eli ratkaisujoukko, jolla pyritään vastaamaan tosielämän ongelmaan. Tässä tutkimusprojektissa konstruktion rakentamisen lähtökohtana oli avoimen lähdekoodin ohjelmistokirjastojen ja neuroverkkojen soveltaminen.

Mallin kehittyessä sen ominaisuuksien testaamisessa hyödynnettiin RGCE Cyber Range –ympäristöä, erityisesti keväällä 2018 RGCE-ympäristössä järjestettyä Puolustusministeriön ja Turvallisuuskomitean johtamaa kansallista kyberturvallisuusharjoitusta KYHA 2018 (https://www.defmin.fi/ajankohtaista/tiedotteet?9_m=9314) ja sen verkkoliikennetaltiota. Näin malin kehitystä ja testaamista varten saatiin modernia ja realistista hyökkäysliikennettä, sekä monipuolisia verkkotopologioita. Tämän avulla pyrittiin välttämään konstruktiiviselle tutkimukselle tyypillinen objektiivisuuden puute. Lisäksi hankkeessa kehitettiin organisaation lähiverkkoa mallintava erillisympäristö osaksi RGCE Cyber Range -ympäristöä, jonka avulla kyettiin luomaan kehitystyössä tarvittavaa verkkoliikennedataa.

IDS-järjestelmäkehityksessä verkkoliikennetaltio on äärimmäisen tärkeässä asemassa. Mallin opettamista varten tulee olla saatavilla monipuolista verkkoliikennedataa, jossa ei ole hyökkäys- tai tunkeutumisliikennettä. Tämän lisäksi mallin testausta varten tarvitaan testiliikennettä, jossa on mukana myös modernia ja monipuolista tunkeutumis- ja haittaohjelmaliikennettä.

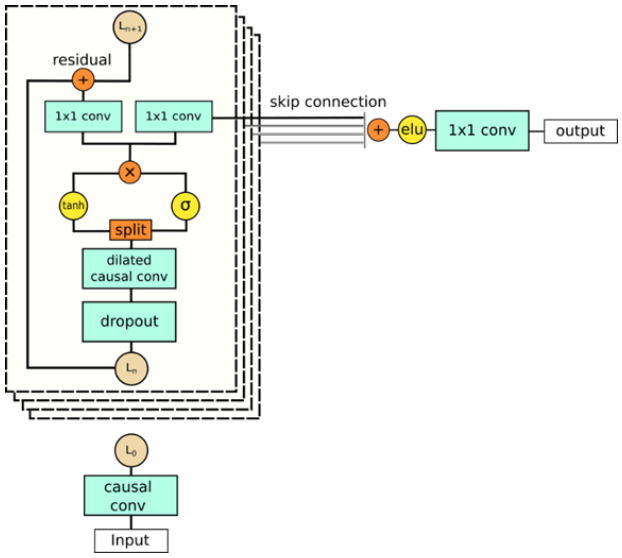
Koska useissa kansainvälisissä tutkimusjulkaisuissa käytetään julkisesti saatavaa testidataa, tulosten vertailun mahdollistamiseksi mallia testattiin myös yleisesti käytetyllä julkisella testidatalla: University of New Brunswick, Canadian Institute for Cybersecurity: Intrusion Detection Evaluation Dataset (CICIDS2017) https://www.unb.ca/cic/datasets/ids-2017.html.

Koska CICIDS2017 datassa ei ollut riittävästi TLS-salattua haittaliikennettä, täydensimme sitä luomalla RGCE-ympäristössä haittaliikennettä käyttäen seuraavia työkaluja: Empire PowerShell post-exploitation agent https://www.powershellempire.com/ sekä Cobalt Strike https://www.cobaltstrike.com/.

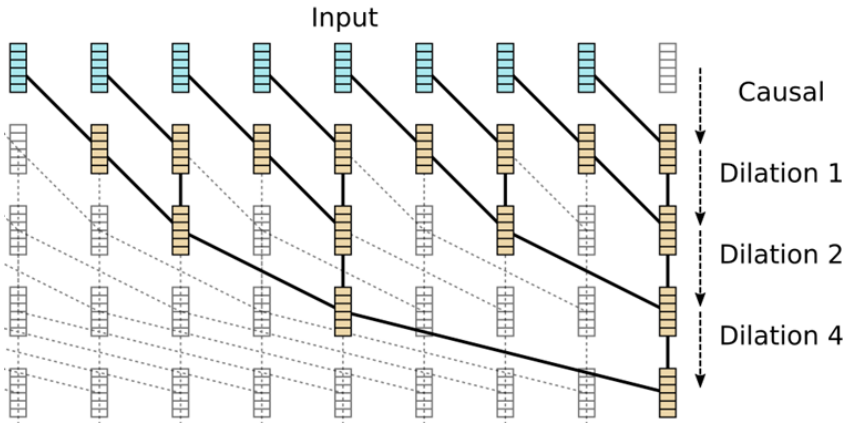
Kehitetyn mallin osalta tavoitteena oli kyetä tunnistamaan poikkeamia TLS-salatusta verkkoliikenteestä. Perusteena TLS-salatun liikenteen analysoinnille on, että salaamaton Internet on käytännössä katoamassa. Modernit tietoverkkoprotokollat hyödyntävät salausta kattavasti. Usein myös haittaohjelmat hyödyntävät komentokanavanaan samaa salattua ja yleisesti käytettyä verkkoliikennekanavaa, jolloin salaus rajoittaa näkyvyyttä analyysimenetelmissä. Toimivan menetelmän on käytännössä kyettävä tunnistamaan anomaliat ilman salauksen purkamista.

Tutkimuksessa parannettiin vuoden 2018 anomaliantunnistusmallia, erityisesti aikasarjoihin liittyvän suorituskyvyn osalta. Vuoden 2018 adversarial autoencoder (AAE) – pohjainen malli (Makhzani et al., 2016, Adversarial Autoencoder, https://arxiv.org/pdf/1511.05644.pdf) vaihdettiin WaveNet (van ven Oord et al., 2016, WaveNet: a Generative Model for Raw Audio, https://arxiv.org/pdf/1609.03499.pdf) pohjaiseen malliin, joka korjaa edeltävän mallin heikkouksia. Molemmat mallit hyödyntävät pakettien ajoituksia verkkoyhteydessä. TLS-yhteyksien vaihteleva pituus ja pakettimäärä luovat lisähaasteita, sillä monet koneoppimisalgoritmit vaativat kiinteänpituisen syötteen.

AAE:in kanssa käytettiin Haar-aallokemuunnosta, jonka avulla verkkoyhteyden pituudesta huolimatta saatiin yhdenpituinen syöte normalisoimalla aallokemuunnoksen tasojen määrä. Autoregressiivisen WaveNetin syöte on teoriassa koko verkkoyhteyden kattava pakettien aikasarja. Syötteeseen voidaan lisätä pakettikohtaisia ominaisuuksia. Verkkopaketeista laskettiin neljä ominaisuutta; paketin suunta, aika seuraavaan lähtevään pakettiin, aika seuraavaan saapuvaan pakettiin sekä paketin koko. WaveNet on alun perin suunniteltu yksiulotteisen äänen mallintamiseen, jonka takia mallin arkkitehtuuriin tehtiin muutoksia (Kuva 1), joista tärkeimmät on 2D-konvoluutio ja häviöfunktion vaihtaminen diskretisoitujen logististen sekajakaumien negatiiviseen uskottavuuteen (Salimans et al., 2017, PixelCNN++: Improving the PixelCNN with Discretized Logistic Mixture Likelihood and Other Modifications. https://arxiv.org/pdf/1701.05517.pdf).



Kuva 1. Wavenet arkkitehtuurikuva. Mallin ”causal conv”-tasot käyttävät 2D-konvoluutiota.



Kuva 2. "Causal"-taso maskaa sisääntulosta ennustettavan paketin ominaisuudet ja laajennetut konvoluutiot kasvattavat verkon näkökenttää eksponentiaalisesti.

Käytännössä pidempien yhteyksien määrä opetusdatassa oli suhteellisen vähäinen, joten WaveNet opetettiin ennustamaan ensimmäiset 256 pakettia kokonaisesta yhteydestä ja lyhyemmät yhteydet täydennettiin täytesymbolilla. Mallin syöte rakennettiin niin, että paketit ovat aikajärjestyksessä vaakasuunnassa, jolloin pakettien ominaisuudet jäävät pystysuuntaan. Malli opetetaan ennustamaan seuraavan paketin ominaisuuksia aikaisempien pakettien perusteella. Tämä on toteutettu maskaamalla konvoluution näkökentästä ennustettava ja sitä seuraavat paketit. Täten malli on täysin riippumaton ajallisesti seuraavista paketeista, mikä mahdollistaa sekä täytesymbolien poislukemisen, että pitkien yhteyksien katkomisen. Verkon näkökenttää suurennettiin käyttämällä laajennettuja konvoluutioita, joiden kernelin solujen välimatkaa toisiinsa nostettiin eksponentiaalisesti tasojen määrän suhteen (Kuva 2).

4. Tulokset ja pohdinta

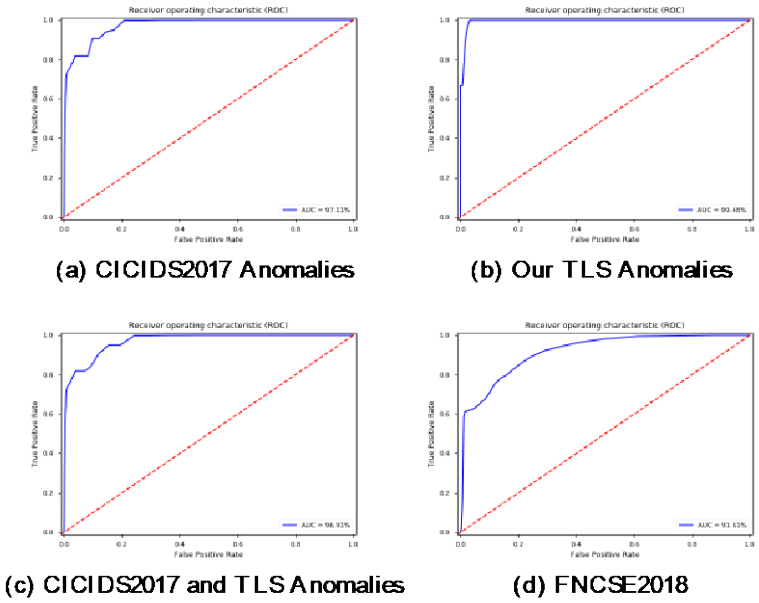
Kehitetyn mallin suorituskykyä on arvioitu Receiver Operating Characteristics (ROC) –kuvaajilla. ROC-kuvaajan alapuolelle jäävää pinta-alaa (Area Under Curve [AUC]) pidetään mallin erottelukyvyn mittarina.

Vertailunkelpoisuuden vuoksi tulokset on laskettu käyttäen eri datasettien sisältöä niin mallin opetukseen kuin testaamiseenkin: Canadian Institute for Cybersecurity, Intrusion Detection Evaluation Dataset (CICIDS2017) ja Finnish National Cyber Security Exercise data (FNCSE2018). Lisäksi CICIDS2017 datasettiin on generoitu omaa haittaliikennettä.

Kaikkien testikäytössä olleiden datasettien tulosten perusteella (Kuva 3 ja Kuva 4) voidaan sanoa, että malli on toimiva ja tehokkaampi kuin edellisvuonna kehitetty malli. Lisäksi, koska tulokset on ajettu puhtaasti käyttäen kansainvälistä referenssidataa, mallia voidaan vertailla kansainvälisesti julkaistuihin IDS tutkimustuloksiin. Myös tässä valossa kehittämämme malli on varsin tehokas tunkeutumisten havainnoinnissa.

Training dataset	Evaluation dataset	AUC
CICIDS2017	CICIDS2017	97.11%
CICIDS2017	Our TLS anomalies	99.48%
CICIDS2017	CICIDS2017 + Our TLS anomalies	96.81%
FNCSE2018	FNCSE2018	91.61%

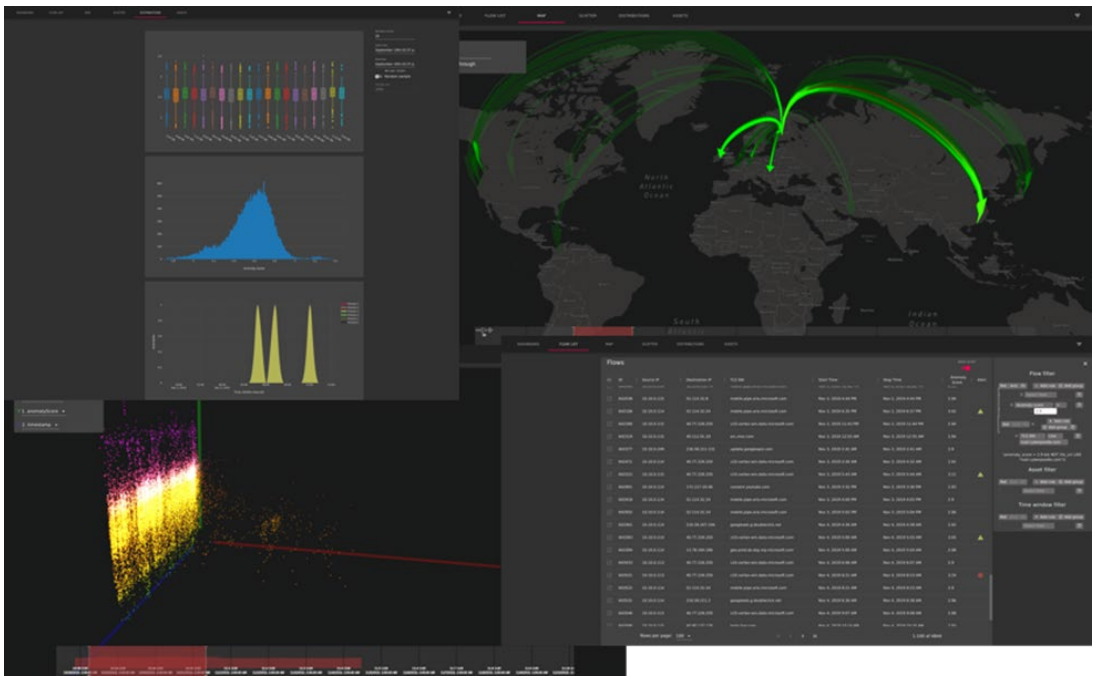
Kuva 3. Havainnointikyvykkyys



Kuva 4. Receiver Operating Characteristics –kuvaajat

Visualisointi toteutettiin käyttäen seuraavia avoimen lähdekoodin työkaluja ja ohjelmistoja: PostgreSQL, Hasura, React, Apollo, Material-UI ja Plotly.js.

Visualisoinnin avulla kyetään parantamaan huomattavasti IDS-mallista saatavien tulosten ymmärrettävyyttä. Ratkaisu parantaa tilannetietoisuutta sekä mahdollistaa paremman päätöksenteon tulosten pohjalta (Kuva 5).



Kuva 5. Visualisoinnin ominaisuuksia

Mallin visualisointia varten kehitetty käyttöliittymä on osoittautunut tehokkaaksi tavaksi parantaa mallista saatavien tulosten ymmärrettävyyttä. Ymmärrettävyys ja tulkittavuus ovat kriittisessä osassa, kun järjestelmää käytetään tilannekuvan muodostamisen ja päätöksenteon pohjana.

Näiden lisäksi tärkeä konkreettinen tulos (joka osoitettiin jo vuoden 2018 tutkimusosuudessa) on, että Open Source-komponentteja käyttäen saadaan kehitettyä tehokas ja moderni koneoppimiseen perustuva tunkeutumisten havainnointijärjestelmä (IDS-järjestelmä).

Tässä yhteydessä täytyy myös mainita edellisvuoden tutkimuksessa kehitetty ympäristö, jossa koneoppimismallia ja datan esikäsittelyjärjestelmää päästiin testaamaan reaaliajassa. Ympäristö mallinsi pientä toimistoverkkoa, jossa automatisoidut botit generoivat verkkoliikennettä ohjaamalla tavallisia työpöytäsovelluksia, kuten selainta, tekstinkäsittelyohjelmaa ja sähköpostiohjelmaa, konenäön avulla. Kehitettyä testiympäristöä voitiin hyödyntää myös tässä tutkimuksessa.

5. Tutkimuksen tuottamat tieteelliset julkaisut ja muut mahdolliset raportit

Tutkimustuloksista on tehty kaksi julkaisua:

- Kokkonen, T., Puuska, S., Alatalo, J., Heilimo, E., Mäkelä, A., ”Network Anomaly Detection Based on WaveNet”, In: Galinina O., Andreev S., Balandin S., Koucheryavy Y. (eds) Internet of Things, Smart Spaces, and Next Generation Networks and Systems. NEW2AN 2019, ruSMART 2019. Lecture Notes in Computer Science, vol 11660. Springer, Cham
- Puuska, S., Kokkonen, T., Mutka, P., Alatalo, J., Heilimo, E., Mäkelä, A., ”Statistical Evaluation of Artificial Intelligence -based Intrusion Detection System”, Submitted for 8th World Conference on Information Systems and Technologies. If accepted, will be published in Advances in Intelligent Systems and Computing, Springer

6. Hankkeen seuraajan lausunto raportista

Lausunto on pyydetty kahdelta oleelliselta seuraajataholta Puolustusvoimilta ja Puolustusministeriöltä. Puoliustusministeriön lausunto on julkinen ja esitetään liitteenä 1.

Molempien lausuntojen osalta voidaan mainita, että tulokset ovat lupaavia, tehty tutkimus nähdään hyödylliseksi kansallisen kyberresilienssin kehittämisen kannalta. Myös jatkotutkimusta aiheen parissa suositellaan.

Liite 1: Hankkeen seuraajan lausunto

Puolustusministeriö
Försvarsministeriet
Ministry of Defence

Lausunto 1 (1)

Harri Mäntylä 31.10.2019 VN/11547/2019
VN/11547/2019-PLM-2

PLM tietohallintoyksikön lausunto Jyväskylän ammattikorkeakoulun MATINE-rahoitteisesta tutkimushankkeesta

Jyväskylän ammattikorkeakoulun IT-instituutissa on vuonna 2019 jatkettu MATINE-rahoitteista tutkimushanketta aiheesta ”Tekoälyn käyttö poikkeamapohjaiseen tunkeutumisten havainnointiin verkkoliikenteestä”. Tutkimushankkeessa on jatkokehitetty tekoälysovellusta, jolla tunnistetaan hyökkäysliikennettä verkkoliikenteen seasta. Hankkeessa kehitetystä mallista on kirjoitettu kaksi tieteellistä julkaisua, joista toinen julkaistiin Springerin Lecture Notes in Computer Science – kokonaisuudessa ja toinen on arvioitavana tällä hetkellä. Lisäksi hankkeessa on kehitetty haitallisen verkkoliikenteen visualisointia osana tilannekuvan muodostamista.

Puolustusministeriön tietohallintoyksikkö on tutustunut JAMK:n tutkimushankkeeseen ja pitää sitä onnistuneena. Hankkeessa toteutettu sensorikehitys ja haitallisen liikenteen visualisointiratkaisu kehittävät kansallisen kyberpuolustuksen kannalta oleellista osaamista ja hanke on puolustusministeriön julkaisemien kyberpuolustuksen kehittämisen strategisten linjausten mukainen.

Tietohallintojohtaja Teemu Anttila

Tietoturvapäällikkö Harri Mäntylä

Jakelu Jyväskylän ammattikorkeakoulu

Tiedoksi Jyväskylän ammattikorkeakoulu, Tero Kokkonen

Postiosoite
Postadress
Postal Address
Puolustusministeriö
PL 31
FI-00131 Helsinki
Finland

Käyntiosoite
Besöksadress
Office
Eteläinen Makasiinikatu 8
00130 Helsinki
Finland

Puhelin
Telefon
Telephone
0295 16001
Internat. +358 295 16001

Faksi
Fax
Fax

s-posti, internet
e-post, internet
e-mail, internet
kirjaamo@defmin.fi
www.defmin.fi