2018 / 2500M-0096

ISSN 1797-3457 (verkkojulkaisu)
ISBN (PDF) 978-951-663-046-8

TIIVISTELMÄRAPORTTI

Tekoälyn käyttö poikkeamapohjaiseen tunkeutumisten havainnointiin verkkoliikenteestä

Tero Kokkonen, Jyväskylän ammattikorkeakoulu IT-instituutti, tero.kokkonen@jamk.fi
Mika Rantonen, Jyväskylän ammattikorkeakoulu IT-instituutti, mika.rantonen@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

Tiivistelmä. Kiihtyvän digitalisaation sekä verkotettujen tietojärjestelmien määrän kasvun seurauksena modernin yhteiskuntamme toiminta on riippuvainen turvallisista tietoverkoista. Ilmiö koskettaa erityisesti Puolustusvoimia ja muita turvallisuusviranomaisia, joiden johtamisjärjestelmät rakentuvat useista tietoverkoista ja -järjestelmistä. Näihin nk. kriittisiin järjestelmiin kohdistuvat kyberhyökkäykset ovat yleistyneet ja monimutkaistuneet kasvattaen tarvetta tunkeutumisten havainnointijärjestelmien kehitykselle. Tutkimuksen tavoitteena oli vertailla, tutkia ja soveltaa uusimpia avoimen lähdekoodin tekoäly- ja syväoppimisohjelmistokirjastoja ja menetelmiä poikkeamapohjaiseen tunkeutumisten havainnointiin verkkoliikenteestä.

Tutkimushankkeen lopputuloksena on kehitetty tekoälypohjainen poikkeamien havainnointijärjestelmä, jolla tietoverkon kautta tapahtuva tunkeutuminen havaitaan. Järjestelmän toiminta testattiin käyttämällä Kansallisen Kyberharjoituksen (KYHA18) aikana luotua tietoliikennekaappausta. Tämän lisäksi tutkimuksessa käytettiin konenäkömenetelmiin perustuvaa keinotekoista käyttäjädataa. Tulosten pohjalta on tehty kansainvälinen tieteellinen julkaisu sekä insinööriopinnäytetyö.

1. Johdanto

Suomen kyberturvallisuusstrategian keskeisiä tavoitteita on kokonaisvaltainen kyky havaita kyberuhat ja suojautua niiltä. Tämän tavoitteen saavuttamiseksi sensorikyvyn kehittäminen on välttämätöntä.

Esimerkiksi Suojelupoliisin Vuosikirjassa (2016) mainitaan verkkovakoiluhavaintojen lisääntyneen viime vuoden aikana. Suojelupoliisin päällikkö Antti Pelttari totesi kirjan julkaisun yhteydessä seuraavasti: "Erilaiset vaikuttamis- ja hybridioperaatiot, informaatiovaikuttaminen ja vakoilu tietoverkoissa ovat avanneet aivan uuden ulottuvuuden, missä ennalta-arvattavaa on yhä vähemmän".

Termillä "tunkeutuminen" (intrusion) tarkoitetaan tekoa tai menetelmää, jolla vaarannetaan tietojärjestelmä tai –verkko. Tunkeutumisen havainnointijärjestelmät (IDS) jaetaan verkkopohjaisiin (Network Intrusion Detection System, NIDS) ja laitekohtaisiin (Host Intrusion Detection System, HIDS) lähestymistapoihin. Verkkopohjaiset IDS –järjestelmät jaetaan kahteen eri kategoriaan niiden toiminnan mukaan; poikkeamien havainnointiin perustuviin havainnointijärjestelmiin (Anomaly Based IDS) ja väärinkäytösten havainnointiin perustuviin havainnointijärjestelmiin (Misuse Based IDS). Poikkeamien havainnointiin perustuville järjestelmille opetetaan verkkoliikenteen normaalimalli, jonka pohjalta järjestelmät pyrkivät havainnoimaan poikkeamia tästä normaalimallista. Poikkeamien havainnointiin perustuvien järjestelmien hyvänä puolena on se, että niillä kyetään löytämään ennalta

tuntemattomia hyökkäyksiä ns. nollapäivähyökkäyksiä, mutta yleensä ne tuottavat suuren määrän vääriä hälytyksiä (False Positive). Väärinkäytösten havainnointiin perustuvat järjestelmät taas etsivät verkkoliikenteestä väärinkäytösten mukaisia tunnettuja kuvioita. Väärinkäytösten havainnointiin perustuvien järjestelmien hyvänä puolena on niiden hyvyys ja tarkkuus havaittaessa tunnettuja hyökkäyksiä, mutta heikkoutena on se, että ne eivät tunnista ennalta tuntemattomia hyökkäyksiä ja toisaalta puutteet tai virheet tunnettujen väärinkäytöskuvioiden määrittelyssä heikentävät havaintokykyä.

Kaikessa turvallisuustoimintaan liittyvässä päätöksenteossa tilannekuvalla ja tilannekuvan perusteella saavutettavalla tilannetietoisuudella on erittäin suuri arvo. Tilannekuvatutkimuksen pioneeri Mica Endlsey on todennut, että jopa koulutettu päätöksentekijä tekee vääriä/virheellisiä päätöksiä huonon tai virheellisen tilannekuvan pohjalta. Näin ollen tilannekuvan pohjana oleva sensoritieto on avainasemassa erityisesti puhuttaessa kyberturvallisuudesta, jossa ilmiöiden havainnollistaminen voi olla kompleksisempaa kuin reaalimaailmassa.

2. Tutkimuksen tavoite ja suunnitelma

Tässä tutkimuksessa sovellettiin syväoppimista tunnistamaan poikkeamia verkkoliikenteestä. Kehitetylle sovellukselle opetettiin normaalin/laillisen verkkoliikenteen malli, jonka perusteella tunnistamisvaiheessa tekoäly pyrkii havaitsemaan poikkeamat opitusta normaalimallista. Sovelluksen avulla verkkoliikennettä voidaan monitoroida jatkuvasti ja opetettu sovellus kykenee reagoimaan poikkeamiin lähes reaaliajassa. Tunnistamista voidaan soveltaa joko koko verkkosegmenttiin tai jopa yksittäiselle käyttäjälle tai koneelle.

Tutkimuksessa tavoitteena oli lisäksi vertailla, tutkia ja soveltaa uusimpia avoimen lähdekoodin (Open Source) tekoälyohjelmistokirjastoja ja muita Open Source ohjelmistokomponentteja mallin implementoimiseen. Näin saadaan samalla tutkittua Open Source komponenttien soveltuvuus tekoälyyn perustuvan reaaliaikaisen sensorimallin kehitykseen.

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 ja erityisesti keväällä 2018 RGCE-ympäristössä järjestettyä Puolustusministeriön ja Turvallisuuskomitean johtamaa kansallista kyberturvallisuusharjoitusta ja sen verkkoliikennetaltiota (KYHA 2018, https://www.defmin.fi/ajankohtaista/tiedotteet?9_m=9314). 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. Kehitystä 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.

Tutkimuksessa käytetty data sekä innovatiivinen koneoppimismalli tekevät siitä maailmanlaajuisestikin merkityksellisen. Useissa tutkimusjulkaisuissa käytetään julkisesti saatavaa testidataa, joka on pääsääntöisesti teknisesti vanhentunutta, liiaksi anonymisoitua sekä kovin yksinkertaista. Tässä tutkimuksessa kyettiin hyödyntämään modernia ja monipuolista kansallisessa kyberturvallisuusharjoituksessa syntynyttä aineistoa, jolloin menetelmän soveltuvuutta moderneihin kyberuhkiin voitiin arvioida tehokkaasti.

Projektissa toteutettiin lisäksi ympäristö, jossa kehitettyä koneoppimismallia ja datan esikäsittelyjärjestelmää voitiin testata reaaliajassa. Ympäristö mallinsi pientä toimistoverkkoa, jossa automatisoidut botit tuottavat verkkoliikennettä ohjaamalla konenäön avulla tavallisia työpöytäsovelluksia, kuten selainta, tekstinkäsittelyohjelmaa ja sähköpostiohjelmaa. Normaalien työpöytäsovellusten käyttö verkkoliikenteen tuottamiseksi on tärkeää, jotta verkkoliikenne on realistista. Työpöytäsovellukset luovat monimutkaista verkkoliikennettä, jonka simuloiminen ohjelmallisesti on erittäin hankalaa. Mikäli liikenne olisi luotu simulaattorilla sen ominaisuudet eivät olisi vastanneet todellisuutta tutkimuksen vaatimalla tarkkuudella. Simuloitu yritysverkko rakennettiin hyödyntäen RGCE-ympäristöä, jolloin haittaliikenteen tuottaminen kyberhyökkäystyökaluilla oli mahdollista. Ympäristöä hyödynnettiin reaaliaikatestauksen lisäksi myös lisädatan generointiin koneälymallin testausta varten.

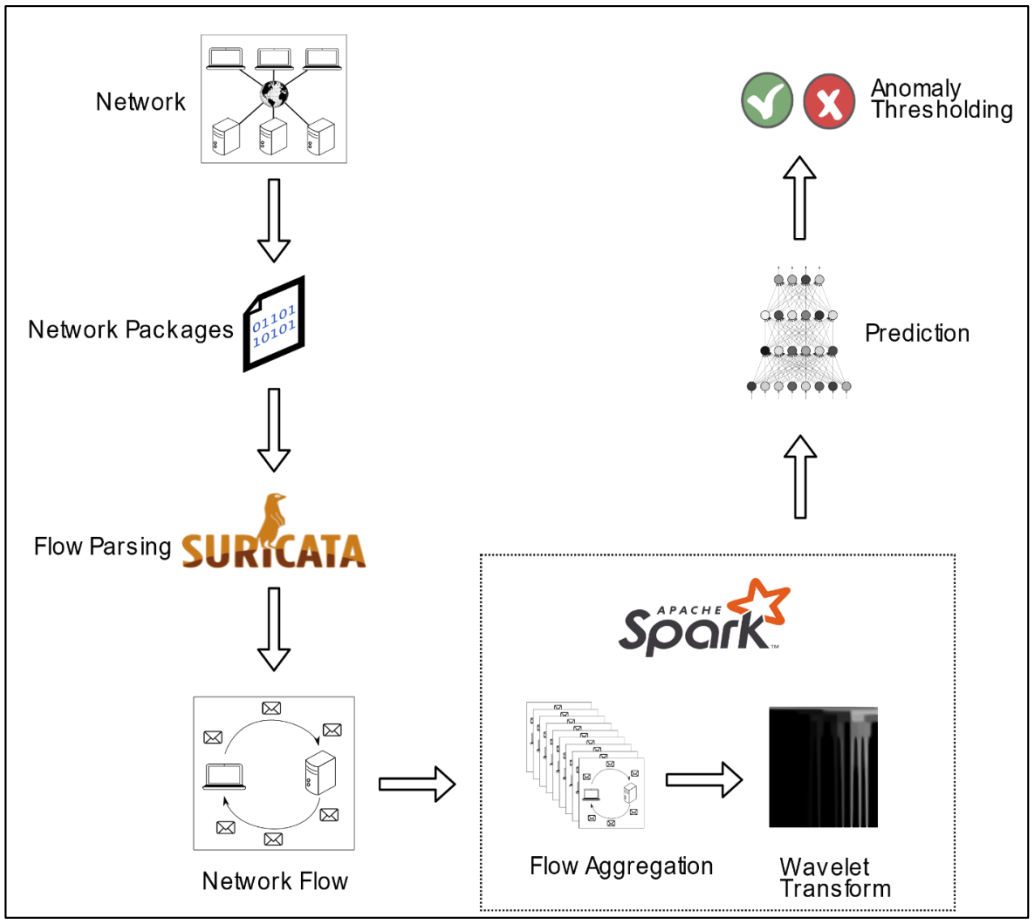
Tutkimuksen tavoitteena mainittu reaaliaikaisuus saavutettiin toteuttamalla verkkoliikennedatan esikäsittelyketju käyttäen moderneja avoimen lähdekoodin komponentteja. Datan esikäsittelyketju (*Kuva 1*) suunniteltiin niin, että se voidaan ottaa helposti käyttöön osaksi realistisia verkkoratkaisuja. Järjestelmän ainoa vaatimus on, että tarkkailtava verkkoliikenne voidaan peilata järjestelmälle jossa on verkkoliikenteen parsimiseen tarkoitettu Suricata NMS ohjelma. Kyseiseen Suricata NMS ohjelmaan tehtiin muutos, joka mahdollisti verkkoliikenteen yksittäisten pakettien ajoitusten keräämisen verkkoliikenneyhteyksistä. Suricata NMS ohjelmaa siis käytettiin pelkästään verkkoliikenteen jäsentelemiseksi erillisiin verkkoliikenneyhteyksiin (network flow), jotta datan käsittely myöhemmissä esikäsittelyketjun vaiheissa olisi helppoa.

Datan esikäsittelyketjun komponenttien välillä tapahtuva tiedonsiirto toteutettiin käyttäen Apache Kafka stream-processing -alustaa. Alusta on erityisesti reaaliaikatietovirtojen (real-time dataflow) hallintaan käytetty ohjelmisto, joka skaalautuu suuriinkin datamääriin. Suricata NMS ohjelma konfiguroitiin lähettämään jäsennelty verkkoliikenneyhteysdata Kafka-alustalle. Muut esikäsittelyketjun komponentit pystyivät lukemaan tätä tietovirtaa ja tekemään myös itse uusia tietovirtoja alustalle.

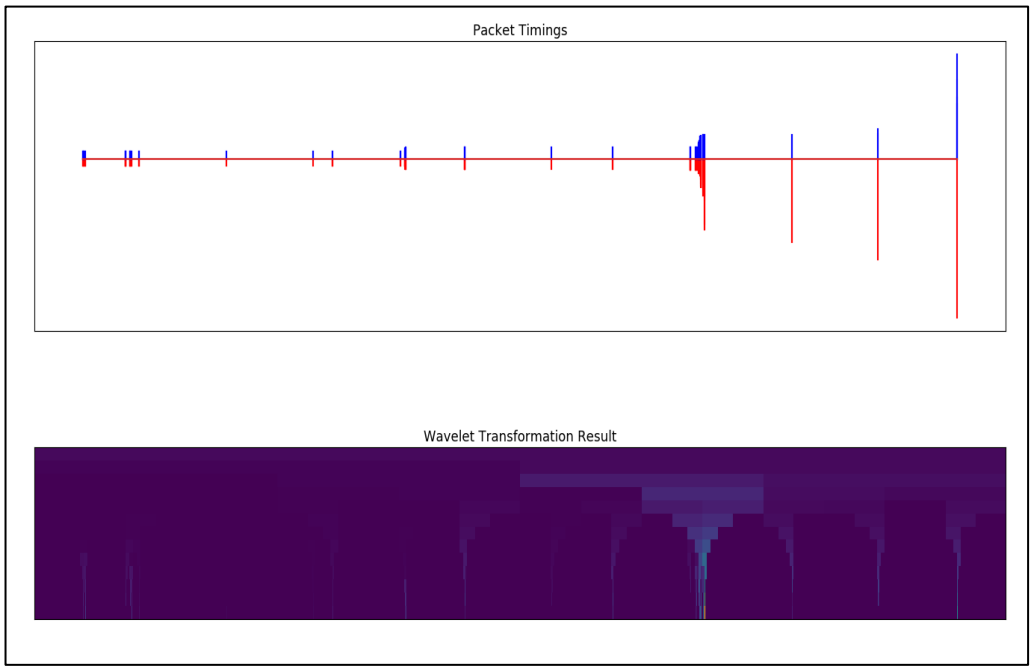
Piirreirrotus (feature extraction) toteutettiin käyttäen Apache Spark klusterilaskentaohjelmistokehystä (cluster-computing framework). Apache Spark on erityisesti tarkoitettu nk. big datan prosessointiin vikasietoisessa klusteroidussa ympäristössä, mutta kyseisellä ohjelmistokehyksellä on mahdollista toteuttaa ratkaisuja viestivirtojen prosessointiin. Järjestelmässä Apache Spark ohjelmistokehyksellä toteutettu piirreirrotusohjelma prosessoi Suricata NMS ohjelman lähettämiä verkkoliikenneyhteyksiä Apache Kafka alustan viestivirrasta. Ohjelma yhdistää määrätyllä aikaikkunalla ja ominaisuuksilla varustettuja verkkoliikenneyhteyksiä viestivirrasta ja laskee aallokemuunnoksen (wavelet transform) verkkoliikenneyhteyden pakettiajoituksista luodusta impulssisignaalista. Lopullisessa toteutuksessa verkkoliikenneyhteyksistä ei käytetty muita ominaisuuksia kuin pelkästään pakettiajoituksia, mutta järjestelmällä on mahdollista irrottaa myös muita piirteitä verkkoliikenneyhteyksistä, jolloin dataa on mahdollista rikastaa muiden menetelmien tuottamalla tiedolla.

Aallokemuunnoksen tarkoituksena on muovata vaihtelevan pituiset verkkoliikenneyhteydet muotoon joka voidaan syöttää tekoälyalgoritmille. Aallokemuunnos laskettiin signaalista, joka generoitiin summaamalla yksittäisten pakettien ajoituksesta ja koosta luodut impulssisignaalit yhteen. Lähtevät ja tulevat paketit yhdistettiin yhdeksi signaaliksi siten, että tulevilla paketeilla oli negatiiviset impulssit ja lähtevillä paketeilla positiiviset (*Kuva 2*). Tut-

kimuksen alussa tarkasteltiin myös Fourier-muunnoksen käyttöä ajoituskuvioiden esiin tuomiseen, mutta pakettiajoituksista saatava signaalin on luonteeltaan erittäin harvaa (sparse signal). Tutkimuksessa huomattiin, että Fourier-muunnoksen käyttö tällaiseen dataan ei tuottanut piirteitä, jotka tutkimuksessa käytetty tekoälyalgoritmi olisi löytänyt. Lisäksi aikasarjat olivat ei-stationaarisia, eikä niitä voi helposti esittää menetelmillä, joissa varianssille on asetettu vaatimuksia.



Kuva 1. Datan esikäsittelyketju



Kuva 2. Esimerkki aallokemuunnoksesta

Tutkimuksessa kehitetty koneoppimismalli perustuu nk. adversariaaliseen autoencoderiin, AAE (Makhzani et al., 2016, Adversarial Autoencoder, https://arxiv.org/pdf/1511.05644.pdf).

AAE:lla on kyky löytää piileviä ominaisuuksia datasta x ja klusteroida se löydettyjen ominaisuuksien perusteella täysin itsenäisesti (unsupervised learning). AAE:n toiminta perustuu perinteiseen autoencoderiin (AE). Perinteinen autoencoder on neuroverkko, jonka tavoitteena on oppia datasta piirteitä itsenäisesti. AE tiivistää sisään tulevan datan matalaulotteisempaan latenttiavaruuteen z (latent space), jonka jälkeen se puretaan tavoitteena saavuttaa alkuperäistä dataa muistuttava esitys. Tällä menettelyllä AE pystyy vähentämään kohinaa ja havaitsemaan datasta yhdistäviä tekijöitä. Neuroverkko optimoidaan siten, että että sisääntulo ja ulostulo $\hat{x}$ vastaavat mahdollisimman paljon toisiaan minimoimalla keskineliövirhe:

$$L_{AE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2$$

AAE hyödyntää nk. generatiivisia adversariaalisia neuroverkkoja, GAN (Goodfellow et al, 2014, Generative Adversarial Networks, https://arxiv.org/pdf/1406.2661.pdf) latenttiavaruuden jakauman $q(z)$ muokkaamiseen jakauman $p(z)$ suuntaiseksi. Tällä saavutetaan latentin avaruuden tehokkaampi käyttö.

GAN sisältää kaksi neuroverkkoa; generaattorin G ja diskriminaattorin D. Generaattorin tarkoitus on luoda dataa, joka matkii oikeaa dataa, kun taas diskriminaattori pyrkii erottamaan generoidun ja oikean datan. Vastakkain asettelemalla diskriminaattori ja generaattori siten, että diskriminaattori maksimoi häviön

$$L_D = \log(D(p(z))) + \log(1 - D(G(x_G)))$$

ja generaattori pyrkii minimoimaan häviön

$$\log(1 - D(G(x))).$$

Täten generaattori oppii huijaamaan diskriminaattoria luomalla alkuperäistä dataa muistuttavaa dataa.

AAE:ssa Koodaaja $q(z|x)$ (encoder) toimii generaattorina, jolloin generaattorin minimoitava häviö on $L_G = \log\left(1 - D\left(q(z)\right)\right)$. Diskriminaattori on erillinen monikerroksinen perceptron-verkko, jota opetetaan, AE:n ja generaattorin kanssa vuorotellen erillisillä optimisaattoreilla.

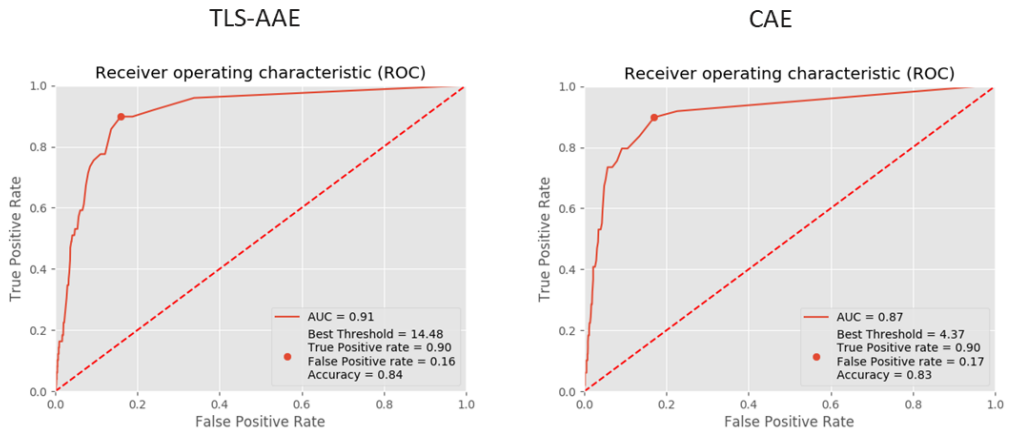
AAE:lla klusterointi tapahtuu jakamalla koodaaja jatkuvaan sekä diskreettiin ulostuloon $q(z, y|x)$. Diskreetin ulostulon y aktivoidaan softmax-funktiolla ja jakauma pakotetaan kategoriseen distribuutioon $y \sim Cat(c)$, jossa c on haluttujen klustereiden määrä. Optimointi tapahtuu lisäämällä verkkoon diskriminaattori, jolle syötettävä oikea jakauma $p(z) = Cat(c)$. Yhtä lailla jatkuva ulostulon z pakotetaan normaalijakaumaan, jonka keskiarvo on nolla $z \sim N(0, \sigma^2)$. Jotta klusterien etäisyyksiä toisiinsa voidaan optimoida, verkkoon lisätään yksikerroksinen neuroverkko $H(h|y)$, jolle syötetään klusterin one-hot vektori, jonka päällä oleva indeksi on sama kuin diskreetin ulostulon y suurimman arvon indeksi. Keskipisteiden välinen etäisyys optimoidaan, niin että etäisyys on aina suurempi kuin 3σ, jolloin 99.7% klustereiden pisteistä ei ole päällekkäin toisten klustereiden kanssa, kun jatkuva ja diskreetti ulostulo summataan. Summa syötetään koodauksen purkajalle, jonka ulostulon $\hat{x}$ ja alkuperäisen datan x rekonstruktiovirheen neliö $(x_i - \hat{x}_i)^2$ toimii anomaliapisteytyksenä.

Tällä menettelyllä latentin avaruuden rakennetta voidaan muokata tasaisemmaksi, jolloin havaintokyky paranee. Menetelmä myös vähentää opetusdatan tarvetta, jolloin malli toimii luotettavammin tilanteessa, jossa dataa on vähän tai se on tilastollisesti epäedustavaa.

4. Tulokset ja pohdinta

Saavutetut tulokset vaikuttavat lupaavilta; mallin avulla kyetään tunnistamaan tunkeutumisia verkkoliikenteestä. Alla on taulukoituna tuloksia, joita on myös vertailtu perinteiseen konvolutionaaliseen autoencoderiin. Adversariaalisen autoencoderin etu perinteiseen autoencoderiin nähden on sen generatiivinen ominaisuus, jonka havaittiin vähentävän väärien hälytyksien määrää generoimalla opetettavan datan välimuotoja.

Kehitetyn mallin ja sovelluksen tehokkuutta on kuvattu piirtämällä mittaustulosten perusteella Receiver Operating Characteristics (ROC) -kuvaajat, joissa kuvaajan alapuolelle jäävä pinta-alaa, Area Under Curve (AUC), pidetään mallin erottelukyvyn mittarina. Tämän lisäksi havainnointikykyä kuvaamaan on laskettu True Positive Rate (TPR), False Positiive Rate (FPR) sekä Accuracy -arvot. Vertailun vuoksi tuloksia on verrattu perinteisellä konvolutionaalisella autoencoreilla (Convolutional Autoencoder, CAE) saavutettuihin tuloksiin (*Kuva 3* ja *Kuva 4*). Näiden perusteella voidaan sanoa, että malli näyttää toimivalta ja on tehokkaampi kuin perinteinen CAE –mali. On kuitenkin huomioitava, että FPR on suhteellisen suuri ja vaatii mallin optimointia. Suuri FPR on yleisesti poikkeamapohjaisten IDS-järjestelmien tunnistettu ongelma, ja sen eteen tehdään työtä myös muissa kansainvälisissä tutkimuksissa. Mallin optimoinnilla voidaan kehittää tuloksia myös oikeiden positiivisten havaintojen osalta paremmaksi.



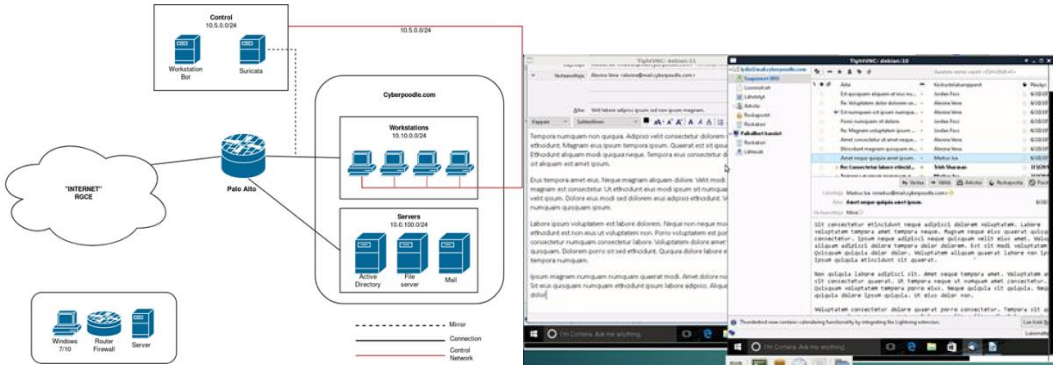
Kuva 3. Receiver Operating Characteristics –kuvaajat

Method	TPR	FPR	Accuracy	AUC
TLS-AAE	0.90	0.16	84%	0.91
CAE	0.90	0.17	83%	0.87

Kuva 4. Havainnointikyvykkyys

Yleisesti ottaen kehitetyssä mallissa on potentiaalia, mutta on myös huomattava, että se vaatii vielä jatkokehitystä ja optimoimista ennen kuin sen perusteella voi esimerkiksi kehittää valmista tuotetta tai ominaisuutta valmiiseen tunkeutumisten tunnistamisjärjestelmään (IDS-järjestelmään).

Näiden lisäksi tärkeä konkreettinen tulos on, että tutkimuksella saatiin todistettua, että Open Source -komponentteja käyttäen saadaan kehitettyä tehokas ja moderni koneoppimiseen perustuva tunkeutumisten havainnointijärjestelmä (IDS-järjestelmä).



Kuva 5. Kehitetty testiympäristö

Tuloksina voidaan myös mainita toteutettu ympäristö, jossa kehitettyä 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 (Kuva 5). Kehitettyä testiympäristöä voidaan hyödyntää myös muissa vastaavanlaisissa tutkimustöissä. Reaaliaikainen järjestelmän testaaminen myös

vahvisti neuroverkosta lasketun tarkkuuden oikeaksi.

5. Loppupäätelmät

Digitalisaation myötä syntynyt yhteiskunnan kokonaisriippuvuus verkotetuista tietojärjestelmistä näkyy myös kasvuna niihin kohdistuneiden hyökkäysten (yleisemmin tunkeutumisten) osalta. Tämä asettaa kasvavia vaatimuksia sensorikyvyn kehittämiselle. Puolustusvoimien tapauksessa erityisesti nykyaikaisen sodan kuvan laajeneminen ja verkottunut puolustusratkaisu kasvattaa verkon kautta tapahtuvia vaikuttamismahdollisuuksia puolustusjärjestelmiä vastaan. Tietoverkoissa tapahtuvien ilmiöiden ymmärtäminen, tilannekuva ja niiden taustalla oleva sensorikyky on tärkeässä asemassa. Tekoälyn voimakas kehittyminen ja tutkimuksessa sovellettavien tekoälykomponenttien käyttö avaavat uusia mahdollisuuksia sensorikyvyn kehittämiselle tunkeutumisten tunnistamiseen verkkoliikenteestä.

Tässä työssä kehitetty kyky antaa hyvän lähtökohdan mallin jatkokehittämiselle ja laajentamiselle. Työssä saavutetut tulokset osoittavat, että kehitetty malli mahdollistaa tekoälyn ja koneoppimisen soveltamisen tunkeutumisten havaitsemiseen verkkoliikenteestä Anomaly Detection -periaatteiden mukaisesti. Jatkokehityskohteita ovat erityisesti mallin optimoiti tehokkuuden kasvattamiseksi, ja toisaalta tulosten visualisoiti tilannekuvan ja päätöksenteon parantamiseksi. Myös raportin liitteenä olevat ”asiakasalausunnot” tukevat tätä näkemystä ja loppupäätelmiä.

Tutkimustulosten pohjalta on julkaistu tieteellinen artikkeli, joka jo nyt on saanut osakseen suurta mielenkiintoa sovelletun mallin uutuuden ja tieteellisen kiinnostavuuden ansiosta. Mahdollisen jatkotutkimuksen tuloksena mallilla voidaan saavuttaa kansainvälisesti uraauurtavia tuloksia.

Tutkimuksella saatiin todennettua Open Source -komponenttien käytettävyys ja soveltuvuus modernin koneoppimiseen perustuvan IDS-järjestelmän kehityksessä. Samassa yhteydessä havaittiin myös se tosiasia, että koneoppimistutkimuksessa käytetyllä datalla on suunnaton merkitys tulosten saavuttamisessa. Tässä yhteydessä saatiin käyttöön kansallisen kyberturvallisuusharjoituksen verkkoliikennetaltiota, joka mahdollisti nykyaikaisten hyökkäysprofiilien ja hyökkäysliikenteen käytön tutkimuksessa ja toisaalta tarpeeksi monimutkaisen verkkotopologian käytön useisiin yksinkertaisiin laboratorioympäristöihin verrattuna. Hankkeen aikana saatiin käytettyä RGCE Cyber Range ympäristöä osana tutkimusta ja toisaalta osaksi RGCE ympäristöä rakennettiin oma organisaation lähiverkkokokonaisuus, joka mahdollistaa automatisoidun käyttäjäverkkoliikennedatan keräämisen. Näitä saadaan hyödynnettyä myös mahdollisessa jatkokehityshankkeessakin.

6. Tutkimuksen tuottamat tieteelliset julkaisut ja muut mahdolliset raportit

Tutkimustuloksista on tehty kaksi julkaisua:

- Puuska S., Kokkonen T., Alatalo J., Heilimo E., "Anomaly-based Network Intrusion Detection using Wavelets and Adversarial Autoencoders", joka on esitelty 8-9.11.2019 International Conference on Information technology and Communications Security SECITC-tapahtumassa (www.secitc.eu). Tutkimus julkaistaan Springerin Lecture Notes in Computer Science -kokonaisuudessa lähiaikoina (lopullisia ISBN-tietoja ei ole vielä saatavilla).
- Heilimo Eppu, Insinööri-opinnäytetyö, "Applying Adversarial Autoencoders to Anomaly Detection in Cyber Security". Työ arvioidaan joulukuussa ja julkaistaan Internetissä Theseus-tietokannassa Jyväskylän ammattikorkeakoulun opinnäytetyönä.

7. Hankkeen seuraajan lausunto raportista

Lausunto pyydetty kahdelta oleelliselta seuraajataholta Puolustusvoimien johtamisjärjestelmäkeskuksen kyberosastolta (Liite1) ja Puolustusministeriön tietohallintoyksiköltä (Liite2).

Molempien lausuntojen osalta mainitaan, että tulokset ovat lupaavia, tehty tutkimus on ansiokas, joka erityisesti nähdään hyödylliseksi kansallisen kyberresilienssin kannalta. Työ vaatii kuitenkin jatkotutkimusta erityisesti mallin jatkokehittämisen osalta.

Puolustusvoimien johtamisjärjestelmä-keskus
Kyberosasto
JYVÄSKYLÄ

Lausunto

AO21184

22.11.2018

Jyväskylän ammattikorkeakoulu
IT-instituutti

Piippukatu 2
40100 JYVÄSKYLÄ

Sähköposti FT Kokkonen JAMK – Evl Kantola PVJJK 26.10.2018

PVJJK:N LAUSUNTO TUTKIMUSHANKKEEN LOPPURAPORTTIIN

1 Pyyntö

Jyväskylän ammattikorkeakoulu (JAMK) on viitteen mukaisesti pyytänyt Puolustusvoimien johtamisjärjestelmäkeskuksen (PVJJK) lausuntoa MATINE-rahoitteisen tutkimushankkeen ”Tekoälyn käyttö poikkeamapohjaiseen tunkeutumisten havainnointiin verkkoliikenteestä” loppuraportista.

2 Lausunto

JAMK on tutkinut haitallisen verkkoliikenteen tunnistamista erityisesti salatun verkkoliikenteen osalta. Verkkoliikenteen siirtyessä yhä enemmän salatuksi aiheuttaa se suuria haasteita haitallisen liikenteen tunnistamiselle. Perinteiset sääntöpohjaiset tunnistamismenetelmät menettävät tehoansa ja uusia menetelmiä tarvitaan havainnointikyvyn takaamiseksi ja parantamiseksi.

PVJJK näkee koneoppimista ja tekoälyä hyödyntävien järjestelmien tulevaisuuden mahdollisuudet hyvin laajoiksi ja aihealueen tutkimuksen tärkeäksi. Sen lisäksi, että järjestelmät mahdollistavat poikkeamien havaitsemisen salatusta liikenteestä, niillä voisi olla mahdollista havaita haitallista liikennettä paljon sääntöpohjaisia järjestelmiä nopeammin, koska tunnistaminen ei perustu tiedossa oleviin havainnointisääntöihin.

Tekoälyllä on mahdollista tehdä havainto organisaation liikenteen poikkeamista ilman muualla tehtyä havaintoa, joka vaatisi tietyn havaintoon perustuvan säännön. Tämä voi lyhentää havainnointiaikaa huomattavasti ja voi mahdollistaa tunkeutujan havaitsemisen paljon nykyistä nopeammin ja tehokkaammin. Tekoälyn käyttäminen havaitsemisessa voi antaa puolustajalle hieman lisää työkaluja järjestelmien suojaamiseen, sillä hyökkääjän on hankalaa tietää kohdeorganisaation liikenneprofiileja ja näin ainakin hyökkäyksen ensivaiheessa haitallisen liikenteen piilottaminen on vaikeampaa.

JAMK:n tutkimuksessa saadut tulokset ovat rohkaisevia, mutta lisätutkimusta aiheesta tarvitaan. Erityisesti havaitsemisen algoritmeja, tehokkuutta ja luotettavuutta tekoälyyn pohjautuvissa järjestelmissä tulee tutkia. Olisi myös kehitettävä menetelmiä kuinka varmistetaan, että kohdejärjestelmä mihin tekoälyyn ja koneoppimiseen perustuva järjestelmä liitetään, on liittämisen hetkellä varmasti puhdas.

Tärkeää olisi myös vertailla perinteisiä sääntöpohjaisten järjestelmien ja tekoälyyn perustuvien järjestelmien tehokkuutta ja eroja keskenään sekä selvittää tarvitaanko perinteisiä järjestelmiä tekoälyä hyödyntävien järjestelmien osana. Väärien havaintojen (ns. False Positive) määrän pienentäminen on hyvin tärkeää järjestelmän luotettavuuden parantamiseksi.

3 Muuta

Lisätietoja lausunnosta antaa ICT-erityisasiantuntija Sami Orasaari puh. 0299 800 (vaihde).

Keskuksen johtaja
Eversti Harri Suni

Osastopäällikkö
Everstiluutnantti Harry Kantola

Tämä asiakirja on sähköisesti allekirjoitettu.

LIITTEET

JAKELU

TIEDOKSI PE JOJÄOS

Puolustusministeriö
Försvarsministeriet
Ministry of Defence

Harri Mäntylä

19.11.2018

VN/6353/2018
VN/6353/2018-PLM-2

PLM tietohallintoyksikön lausunto Jyväskylän ammattikorkeakoulun MATINE-rahoitteisesta tutkimushankkeesta

Jyväskylän ammattikorkeakoulun IT-instituutissa on tehty MATINE-rahoitteista tutkimushanketta aiheesta ”Tekoälyn käyttö poikkeamapohjaiseen tunkeutumisten havainnointiin verkkoliikenteestä”. Tutkimushankkeessa kehitettiin tekoälysovellus, jolla tunnistetaan hyökkäysliikennettä verkkoliikenteen seasta. Hankkeessa kehitetystä mallista on kirjoitettu tieteellinen julkaisu, joka julkaistaan Springerin Lecture Notes in Computer Science –kokonaisuudessa.

Puolustusministeriön tietohallintoyksikkö on tutustunut JAMK:n tutkimushankkeeseen ja pitää sitä onnistuneena. Hankkeessa on kehitetty kansallisen kyberresilienssin kannalta oleellista osaamista ja hankkeessa kehitetty malli vaikuttaa ensimmäisten tulosten perusteella toimivalta.

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