Informaatioteknologian tiedekunnan julkaisuja
No. 15/2014

Martti Lehto, Pekka Neittaanmäki

Kyberturvallisuuden ja big data-analyysin tutkimus ja opetus

JYVÄSKYLÄN YLIOPISTO

Informaatioteknologian tiedekunnan julkaisuja
No. 15/2014

Editor: Pekka Neittaanmäki
Covers: Kati Valpe
Cover picture: Seppo Tarvainen

ISBN 978-951-39-6042-1 (verkkoj.)
ISSN 2323-5004

Jyväskylän yliopistopaino, Jyväskylä 2014

Kyberturvallisuuden ja big data-analyysin tutkimus ja opetus

Martti Lehto, Pekka Neittaanmäki

SISÄLLYS

1 JOHDANTO

Informaatioteknologian tiedekunta vastaa kehittyvän informaatioteknologian sekä digitalisoitumisen tuomiin tutkimus- ja koulutushaasteisiin. Tiedekunta yhdistää kokonaisvaltaisesti teknologian, informaation, organisaatioiden ja liiketoiminnan sekä ihmisen näkökulmat niin tutkimuksessa, koulutuksessa kuin sidosryhmäyhteistyössä. Tiedekunta kouluttaa informaatioteknologian laaja-alaisia ja kansainvälisiä osaajia sekä kauppatieteellisellä että luonnontieteellisellä koulutusalalla.

Informaatioteknologian tiedekunnalla on keskeinen rooli yliopiston painoaloihin kuuluvan ihmisläheisen teknologian kehittämisessä. Tiedekunnan keskeinen vahvuus on kyvykkyys tarkastella informaatioteknologiaa laajasti, useita näkökulmia yhdistäen ja eri ilmiöiden yhteisvaikutuksia tunnistaen. Tämä yhdistyy kansainvälisesti arvostettuun huippututkimukseen kärkialoilla ja aktiiviseen toimijuuteen ympäröivän yhteiskunnan kanssa.

IT-tiedekunta on saavuttanut johtavan aseman laskennallisissa tieteissä, kyberturvallisuudessa, tietojärjestelmätieteissä ja edustaa ainoana IT-alan tiedekuntana kognitiotieteen tutkimusta ja opetusta.

IT-tiedekunnan kansallinen yhteistyö ulottuu useisiin maan yliopistoihin ja ammattikorkeakouluihin. Erityisesti yhteistyö on korostunut kyberturvallisuudessa, jossa yhteistyöverkostoon kuuluu jo 12 yliopistoa ja 5 ammattikorkeakoulua.

Informaatioteknologian tiedekunnassa tehdään kansainvälisesti korkealaatuista IT-alan tutkimusta. Tiedekunnan tutkimushankkeet liittyvät usein yhdessä kansallisten ja kansanvälisten tutkimuskumppaneiden ja teollisuuden kanssa tehtäviin tutkimus- ja kehityshankkeisiin. Rahoittajina ovat Suomen Akatemia, Tekes, EU, säätiöt, yritykset ja monet muut tukijat. Tavoitteena on lisätä erityisesti kansanvälistä rahoitusta kuten Horisontti2020, ERC, ulkomaalaiset säätiöt, yliopistot ja yritykset.

Tiedekunnassa on kaksi opetukseen ja tutkimukseen keskittyvää ainelaitosta: Tietojenkäsittelytieteiden laitos ja Tietotekniikan laitos.

Tietotekniikan laitoksen tutkimus perustuu pääosin analyyttis-konstruktiivisten menetelmien käyttöön teknisestä, laskennallisesta, matemaattisesta tai pedagogisesta näkökulmasta.

Tietojenkäsittelytieteiden laitoksen tutkimuksessa tarkastellaan tietojärjestelmiä ja tietojenkäsittelyä neljästä näkökulmasta: teknologinen, ihmiskeskeinen, liiketoiminnallinen ja informaatiokeskeinen. Nämä näkökulmat muodostavat laitoksen

yleisen tehtävän: ymmärtää, kehittää, suunnitella ja hallita tietojärjestelmiä ja tietojenkäsittelyä sekä niiden vaikutuksia kokonaisvaltaisesti käyttökontekstissaan.

Pekka Ala-Pietilän johtama ICT 2015 -työryhmä esittää tietojenkäsittelyn syväosaamisen kehittämistä: *”Kansainvälisesti kilpailukykyisen ICT-intensiivisen tuotteen ja palvelun kehittämiseen tarvitaan laajaa osaamista. Onnistumisen kannalta on keskeistä, että yrityksellä on käytössään tietotekniikan syväosaajien ydintiimi, joka hallitsee syvällisesti tietojenkäsittelytieteen keskeiset osa-alueet. Kriittisen tärkeitä alueita ovat muun muassa algoritmisuunnittelu, diskreetit rakenteet sekä ohjelmointikielten periaatteet. Samoin tiimillä pitää olla osaamista uusimmista ohjelmistoteknologioista ja kyky soveltaa viimeisintä teknologiaa sovellusten vaatimalla tavalla. Tämä tarkoittaa muun muassa hajautettuja järjestelmiä, verkkoja, tietokantoja, tiedonlouhintaa, koneoppimista, pilvilaskentaa (cloud computing), sulautettuja järjestelmiä, tekoälyä ja kryptologiaa. Tämän vuoksi tarvitaan kansallinen ohjelma vahvistamaan osaamispohjan kehittymistä korkeakouluissa. Ohjelman tulee nopeuttaa yliopistossa ja ammattikorkeakouluissa olevan osaamisen siirtymistä yrityksiin, jotta uusien ICT-intensiivisten tuotteiden ja palvelujen kehittyminen vauhdittuu.”* [2]

IT-tiedekunnan tutkimus ja koulutus on rakennettu vastaamaan kansainvälisiä standardeja, ICT-2023 työryhmän suosituksia ja alan kansallisia ja kansainvälisiä strategioita ja ohjelmia. Tämän lisäksi koulutusohjelmat on rakennettu monitieteellisestä näkökulmasta luomalla laaja kokonaisuus tietojärjestelmätieteistä, tietojenkäsittelytieteestä, laskennallisista tieteistä, sovelletusta matematiikasta, kognitiotieteestä ja koulutusteknologiasta.

IT-tiedekunnan koulutuksessa kandidaattikoulutus antaa perustiedot IT-alasta, maisteriopinnoissa opiskelija voi yksilöllisesti erikoistua oman kiinnostuksensa mukaan ja suorittaa opinnot joko suomen tai englanninkielellä. Tohtorikoulutus toteutetaan englanninkielisissä tutkimusryhmissä kansainvälisessä yhteistyössä.

Maisteriohjelmat jakautuvat viiteen eri ryhmään.

1) Tietotekniikan maisteriohjelmat:
   - ohjelmistotekniikka, tietoliikenne, ohjelmointikielten periaatteet, pelit ja pelillisyys, sensoriverkot, laskennalliset tieteet ja sovellettu matematiikka
   - englanninkielinen: Web Intelligence and Service Engineering (WISE)

2) Tietojärjestelmätieteen ja tietojenkäsittelytieteen maisteriohjelmat:
   - tietojärjestelmätieteen maisteriohjelmassa voi suuntautua tietohallintoon, tietojärjestelmäkehitykseen, sosiaaliseen mediaan sekä käyttäjä- ja ihmislähtöiseen teknologiaan
   - englanninkieliset: Service Innovation and Management (SIM) ja Software Engineering and Service Design -maisteriohjelmat

3) Kognitiotieteen maisteriohjelma: ihmisen ja tietokoneen vuorovaikutus, käytettävyys

4) Informaatioturvallisuus, jossa voi erikoistua joko teknologiaan tai johtamiseen

5) Koulutusteknologia: tietotekniikan aineenopettajankoulutus

2 KYBERTURVALLISUUDEN TUTKIMUS JA OPETUS

2.1 Kyberturvallisuuden osaamisen tarve

Globaali kybertoimintaympäristö muodostuu monimutkaisesta ja -kerroksisista informaatioverkostoista, joihin kuuluu kansallisia julkishallinnon, yritysmaailman ja turvallisuusviranomaisten kommunikaatioverkkoja sekä teollisuuden ja kriittisen infrastruktuurin valvonta- ja ohjausjärjestelmiä, jotka internetin välityksellä muodostavat maailmanlaajuisen verkoston. Tähän globaaliin verkostoon käyttäjät ovat liittyneet erilaisen älykkäiden päätelaitteiden avulla.

Informaatioteknologian vallankumous on kehittänyt Suomea 1990-luvulta alkaen kohti tietoyhteiskuntaa. Muutoksen voimana on ollut kansallinen tietotekniikkaosaaminen ja tehokas telekommunikaatioklusteri. Kansalaisten ja elinkeinoelämän tarpeista lähtevä tiedon monipuolinen jalostaminen ja hyödyntäminen ovat yhteiskunnan tärkeimpiä menestystekijöitä. Tiedosta on tullut yhteiskunnan keskeinen voimavara, jota informaatioteknologian avulla voidaan hyödyntää tehokkaammin kuin koskaan aikaisemmin. Suomi on yksi kehittyneimmistä digitaalisista tietoyhteiskunnista, jonka toiminnat ovat riippuvaisia erilaisista digitaalisista verkoista ja niiden antamista palveluista. Yhteiskunnan kriittinen infrastruktuuri koostuu erilaisista julkisinten ja yksityisten organisaatioiden verkostoista. Tietoteknisten laitteiden ja järjestelmien toimimattomuus, informaatioinfrastruktuurin luhistuminen tai vakavat kyberhyökkäykset voivat aiheuttaa kielteisiä vaikutuksia julkisiin palveluihin, liike-elämään ja hallintoon ja siten koko yhteiskunnan elintärkeisiin toimintoihin.

Suomen kyberturvallisuusstrategian (2013) mukaan *”kyberturvallisuuteen tähtäävän tutkimuksen, kehittämisen ja koulutuksen toteuttaminen eri tasoilla vahvistaa kansallista osaamista ja Suomea tietoyhteiskuntana. Informaatiojärjestelmien toimimattomuus, informaatioinfrastruktuurin luhistuminen tai vakavat kyberhyökkäykset teollisuuden ohjausjärjestelmiä vastaan voivat aiheuttaa kielteisiä vaikutuksia julkisiin palveluihin, liike-elämään ja hallintoon ja siten koko yhteiskunnan toimintaan. Julkishallinto ja elinkeinoelämä tarvitsevat yhä enemmän kyberturvallisuuden erityisosaajia turvaamaan yhteiskunnan ja liiketoiminnan toimivuus.”* [3]

Kyberturvallisuusosaaminen on eri toimintasektoreita poikkileikkaava. Kyberturvallisuuden huippuosaamista tarvitaan, jotta voidaan aikaansaada ja kehittää kybertilannetietoisuutta, tehokasta varautumista kyberuhkatilanteisiin, luoda kriittisiä infrastruktuureita suojaavia järjestelmiä ja kehittää vaikuttavia kyberturvallisuusratkaisuja. Ratkaisujen ja järjestelmien ohella kehitetään osaamista

tilanteiden kokonaisvaltaiseen hallitsemiseen kuten kansalaisten sosiaalisen median turvataitoja, yhteisöviestinnän/kriisiviestinnän työkaluja, riskiryhmien tunnistamista sekä tietoisuuden ja turvataitojen parantamista valistuksen ja kouluopetuksen avulla.

Informaatioturvallisuuden maisteriohjelma perustuu tietojenkäsittelytieteiden ja tietojärjestelmätieteen perustalle. Tietojenkäsittelytiede tieteenalana tutkii tietotekniikkaan, käyttöön liittyviä ongelmia ja kaikkea tietoon liittyviä laskennallisia kysymyksiä. Tietojärjestelmätiede tarkastelee tietojärjestelmien kehittämistä, sovelluksia, käyttöä, johtamista ja vaikutuksia eri konteksteissa. Kyberturvallisuus on näitä tieteenaloja läpileikkaava ja se ulottuu laajaan skaalaan teknologioita ja prosesseja suojattaessa verkkoja, tietokoneita, ohjelmia ja dataa kyberhyökkäyksien vaikutuksilta ja vahingoittumisilta.

2.2 Kyberturvallisuuden tutkimus

2.2.1 Perusteita

Informaatioturvallisuuden maisterikoulutus (INTU) perustuu oman pääaineensa tietojenkäsittelytieteiden perustalle sekä vahvaan Informaatioteknologian tiedekunnassa toteutettavaan tutkimukseen.

Tietojenkäsittelytiede tieteenalana tutkii tietotekniikkaan ja sen käyttöön liittyviä ongelmia. Perinteisessä tietojenkäsittelytieteessä tutkitaan kaikkea tietoon liittyviä laskennallisia kysymyksiä, mutta nykyään tutkimusala on hyvin laaja. Kyberturvallisuus on koko tieteenalaa läpileikkaava ja se ulottuu laajaan skaalaan teknologioita ja prosesseja suojattaessa verkkoja, tietokoneita, ohjelmia, dataa kyberhyökkäyksiltä ja vahingoittumisilta. Osaamistarpeen perusta ulottuu tietojärjestelmätieteeseen, informaatioteknologiaan ja tietojenkäsittelytekniikkaan.

Tietotekniikan laitoksella tutkitaan tietotekniikkaa teknis-matemaattisesta näkökulmasta. Tutkimuksen painoalat liittyvät informaatioteknologian keskeisiin alueisiin, kuten uudenlaisten tietojenkäsittelysovellusten ja ohjelmistojen suunnitteluun, tietoverkkojen tiedonsiirtojärjestelmien suunnitteluun ja hallintaan sekä tehokasta tietokonelaskentaa hyödyntävien numeeristen ja matemaattisten menetelmien ja mallien käyttöön.

Tietojenkäsittelytieteiden laitoksen tutkimuksessa tarkastellaan tietojärjestelmiä ja tietojenkäsittelyä yhdistäen innovatiivisesti ja monitieteisesti neljä keskeistä näkökulmaa: teknologinen, ihmislähtöinen, liiketoiminnallinen ja informaatiolähtöinen. Kyberturvallisuuden tutkimusaloja ovat mm. tieto- ja kyberturvallisuusstrategian kehitysmenetelmät, tietoturvan johtaminen ja hallinta, turvallisten tietojärjestelmien kehitysmenetelmät, tietoturvakäyttäytymisen ja tietoturvakulttuurin parantaminen, tietoturvainvestoinnit sekä Social engineering ja phishing.

2.2.2 Kyberturvallisuus Jyväskylän yliopiston kesäkoulussa 2012

Kesäkoulun 2012 8.-24.8.2012 kurssilla Cyber security – Strategic Power of Future (3 op) käsiteltiin kyberturvallisuutta yhteiskunnan näkökulmasta. Kybermaailman ilmiöitä ja tapahtumia analysoitiin sosiaalisen systeemimallin näkökulmasta ja laajennettiin ymmärrystä inhimillisestä toiminnasta tietoyhteiskunnan kommunikaatioympäristössä. Organisaatioita tarkasteltiin kyberkontekstissa ja pyritään luomaan rationaalinen näkökulma organisaatioiden kykyyn suodattaa oikea tieto väärästä. Lisäksi käsiteltiin kriittisen infrastruktuurin ja SCADA -järjestelmien suojaamista, analysoidaan riskejä ja turvallisuuden toteuttamisen periaatteita.

Luennot:
Professor William Hutchinson, Security Research Centre (SECAU) School of Computer and Security Science, Edith Cowan University, Perth, Australia:

- Deception and the management, security and intelligence functions within organizations

Professor Matthew Warren, School of Information Systems, Faculty of Business and Law, Deakin University, Melbourne, Australia

- Cyber Security in the context of critical infrastructure and SCADA systems

Professor Rauno Kuusisto, University of Jyväskylä, Department of Mathematical information technology, University of Jyväskylä, Finland

- Profiling Cyber World as a Social System

2.2.3 12th European Conference on Information Warfare and Security, ECIW 2013

Jyväskylän yliopiston informaatioteknologian tiedekunta järjesti 11.–12. heinäkuuta 2013 kyberturvallisuuden ja -puolustuksen kansainvälisen konferenssin: 12th European Conference on Information Warfare and Security (ECIW 2013). Konferenssi järjestettiin yhteistyössä Academic Conferences & Publishing Internationalin (Lontoo) kanssa ja se käsitteli laaja-alaisesti kyberturvallisuutta ja kybersodankäyntiä.

Konferenssissa alan huippututkijat ympäri maailman esittelevät laaja-alaisesti kyberturvallisuutta ja kyberpuolustusta käsittelevää tutkimustaan niin tieteellisestä kuin inhimillisestä näkökulmasta. Esiteltäviä tutkimusraportteja on lähes 100 ja osallistuvia tutkijoita on Suomen lisäksi mm. Australiasta, Alankomaista, Tšekin tasavallasta, Virosta, Ranskasta, Ruotsista, Unkarista, Israelista, Italiasta, Irlannista, Portugalista, Kreikasta, Venäjältä, Etelä-Afrikasta, Turkista, Taiwanista, Isosta-Britanniasta ja Yhdysvalloista.

Jyväskylän yliopiston tutkijoiden tutkimusraportteja olivat:

- Amir Averbuch and Pekka Neittaanmäki: Anomaly Detection via Manifold Learning
- Michael Kiperberg and Nezer Zaidenberg: Efficient Remote Authentication

- Jaana Kuula, Olli Kauppinen, Vili Auvinen, Santtu Viitanen, Pauli Kettunen and Tuomo Korhonen: Alerting Security Authorities and Civilians with Smartphones in Acute Situations
- Rauno Kuusisto and Tuija Kuusisto: Strategic Communication for Cyber Security Leadership
- Martti Lehto: The Ways, Means and Ends in Cyber Security Strategies
- Amit Resh and Nezer Zaidenberg: Can Keys be Hidden Inside the CPU on Modern Windows Host
- Alexander Semenov: Analysis of Services in Tor Network: Finnish Segment
- Riku Nykänen and Mikko Hakuli: Information Security Management System Standards: A gap Analysis of the Risk Management in ISO 27001 and KATAKRI

2.2.4 Tutkimus- ja kehityshankkeita

Kyberturvallisuuden tutkimushankkeita ovat:

Professori Pekka Neittaanmäki: New System for Cyber Attacks Protection of Critical Infrastructures 2012–2014, CAP-projekti (1.11.2012–31.10.2014, Tekes):

> CAP-hankkeessa kehitetään innovatiivista tietojärjestelmien turvaamiseen liittyvää menetelmää. Menetelmällä voidaan tunnistaa sellaisetkin järjestelmään kohdistuvat uhat, joita ei aikaisemmin ole voitu tunnistaa ja joista ei ole digitaalista sormenjälkeä kuten tietokoneviruksien jäljittämisessä. Menetelmä tutkii tietomassoista epänormaaleja käyttäytymismalleja ja tekee analyysin pohjalta päätelmiä havaintojen vakavuudesta tietojärjestelmän turvallisuudelle.

Professori Pekka Neittaanmäki: Truly-Protect - Platform for Copyright Protection (1.11.2012–30.10.2014, Tekes):

> Truly Protect -hankkeessa kehitetään digitaalisen median suojausmenetelmää, joka on nykyisin käytössä olevia tehokkaampi ja käytännössä tekee mahdottomaksi digitaalisen median laittoman kopioinnin. Truly Protect soveltuu moniin eri kohteisiin kuten videopelien, elokuvien ja musiikin jakelun turvaamiseen. Menetelmän etuna on se, että se ei vaadi uusia laitteistokomponentteja vaan on täysin ohjelmistopohjainen.

2.3 Kyberturvallisuuden opetus

Informaatioturvallisuuden maisteriohjelma perustuu tietojenkäsittelytieteiden ja tietojärjestelmätieteen perustalle. Tietojenkäsittelytiede tieteenalana tutkii tietotekniikkaan, käyttöön liittyviä ongelmia ja kaikkea tietoon liittyviä laskennallisia kysymyksiä. Tietojärjestelmätiede tarkastelee tietojärjestelmien kehittämistä,

sovelluksia, käyttöä, johtamista ja vaikutuksia eri konteksteissa. Kyberturvallisuus on näitä tieteenaloja läpileikkaava ja se ulottuu laajaan skaalaan teknologioita ja prosesseja suojattaessa verkkoja, tietokoneita, ohjelmia ja dataa kyberhyökkäyksien vaikutuksilta ja vahingoittumisilta.

Informaatioturvallisuuden maisterikoulutus (120 op) on Informaatioteknologian tiedekunnan ainelaitosten yhteinen koulutusohjelma, jonka tavoitteena on luoda opiskelijalle vankka osaaminen työskentelyyn informaatio/kyberturvallisuuden kokonaishallinnan vaativissa johtamis- ja kehittämistehtävissä. Opiskelijalle tarjotaan syventäviä opintoja informaatioturvallisuuden kokonaisuuteen sekä informaatioturvallisuuden eri osa-alueille. Informaatioturvallisuuden opetus muodostuu opintokokonaisuudessa, jossa tarkastellaan kybermaailmaa ja sen turvallisuutta yhteiskunnallisesta, toiminnallisesta, teknologisesta ja systeemisestä näkökulmasta.

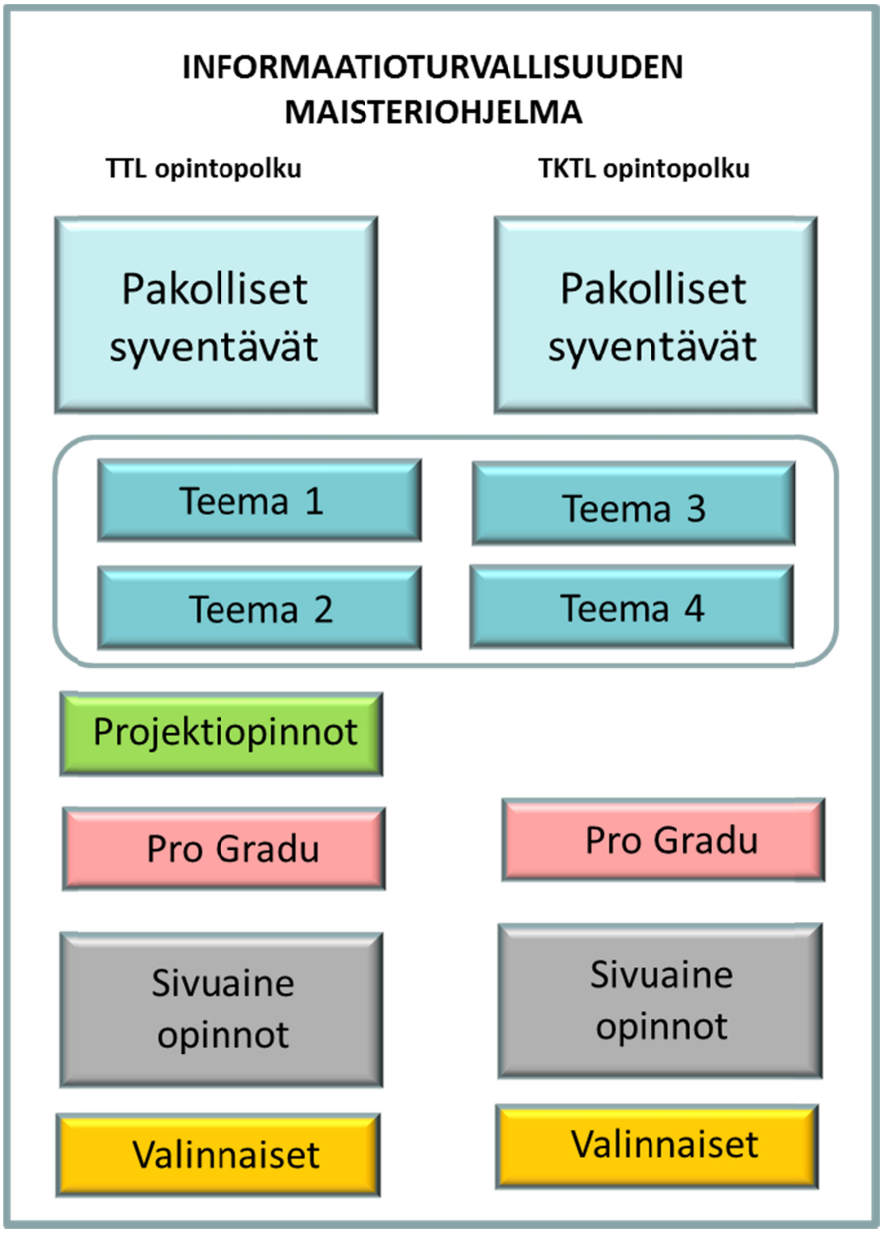
Teknologiaperustaisesta suuntautumisvaihtoehdosta valmistunut maisteri kykenee määrittelemään tietoon, tietoverkkoihin, ja -liikenteeseen sekä tieto- ja ohjausjärjestelmiin sekä toimintaprosesseihin liittyviä informaatioturvallisuusriskejä. Hän tuntee kybermaailman eri uhkamallit ja tuntee uhkien torjuntaan liittyvät toiminnalliset ja teknologiset ratkaisumallit. Hänellä on hyvät valmiudet suunnitella, toimeenpanna ja johtaa informaatioturvallisuuden teknologista suunnittelua ja kehittämistä.

Tähän opintopolkuun kuuluu neljä pakollista kurssia, joita ovat *Kybermaailma ja turvallisuus, Ohjelmistoturvallisuus*, *Tietoverkkoturvallisuus* ja *Requirements Engineering*. Valinnaiset syventävät opinnot on jaettu kolmeen teemaan, joita ovat *Cyber Security Technology, Systems Security* ja *Cyber Conflict*.

Organisaatiolähtöisestä suuntautumisvaihtoehdosta valmistunut maisteri erikoistuu erityisesti tietoturvallisuuden suunnitteluun, johtamiseen ja tietoturvallisuusriskien hallintaan. Hän tuntee kybermaailman eri uhkamallit ja tuntee uhkien torjuntaan liittyvät ratkaisumallit. Hän kykenee johtamaan erilaisten organisaatioiden tietoturvatoimintoja.

Tähän opintopolkuun kuuluu neljä pakollista kurssia, joita ovat *Kybermaailma ja turvallisuus, Information Security Management, Tietoverkkoturvallisuus* ja *Requirements Engineering*. Valinnaiset syventävät opinnot muodostuvat teemakursseista.

Kuvassa 1 on esitetty informaatioturvallisuuden maisteriohjelman rakenne.



KUVA 1 Informaatioturvallisuuden maisteriohjelman rakenne

2.3.1 Kyberturvallisuuden jatkokoulutus

Vuonna 2014 julkaistiin seuraavat kyberturvallisuutta käsitteleviä väitöskirjoja:

- Mikhail Zolotukhin, On data mining applications in mobile networking and network security
- Frans Laakso, Studies on high speed uplink packet access performance enhancements
- Antti Juvonen, Intrusion detection applications using knowledge discovery and data mining

Liitteessä 1 on esitetty kyberturvallisuuden kurssitarjonta lukuvuonna 2014–2015 ja liitteessä 2 on luettelo kyberturvallisuuden jatko-opiskelijoista.

3 BIG DATA-ANALYYSI

3.1 Big data-analyysin osaamisen tarve

ICT 2015 työryhmä toteaa: ”*Digitaalisessa maailmassa informaation ja tallennetun tiedon määrä on valtava. Kun yhdistetään älykkäästi ja reaaliajassa näennäisesti turhaa tietoa, pystytään luomaan täysin uudentyyppistä, toimialojen rajoja rikkovaa tietoa. Big data on maailmalla kuuma tutkimuksen ja soveltamisen kohde. Suomen osaaminen tällä alueella on kapeaa, vaikkakin tietyiltä aloilta löytyy huippuosaamista. Big data liittyy läheisesti muihin Suomen kriittisiin avainosaamisalueisiin. Tietoliikenteen osaajina olemme perinteisesti käsitelleet suuria datamääriä. Tietoturva on tärkeää kaikissa big data -tyyppisissä sovelluksissa ja tietovarannoista louhittavien tiedonjyvästen integrointi vaatii vahvaa ohjelmisto- ja tietojenkäsittelyosaamista. Julkisella puolella tietovarantojen avaaminen, yhteinen ICT -palveluarkkitehtuuri ja perinteisesti Suomessa hyvin toimiva julkinen – yksityinen -yhteistyö avaa mielenkiintoisia mahdollisuuksia. Suurimmaksi ongelmaksi big data tiedon soveltamisessa ja tietojen avaamisessa toimijat kokevat, että organisaatioilla ei ole riittävästi asiantuntemusta. Organisaatiot tarvitsevat tähän osaajia ja koulutusta.”* [2]

Maaliskuussa 2013 Advanced Scientic Computing Advisory Committee (ASCAC) alakomitea julkaisi raportin “Synergistic Challenges in Data-Intensive Science and Exascale Computing”, jossa käsiteltiin “Big Data and the Fourth Paradigm” -teemaa. Raportin mukaan “Historically, the two dominant paradigms for scientific discovery have been theory and experiments, with large-scale computer simulations emerging as the third paradigm in the 20th century. In many cases, large-scale simulations are accompanied by the challenges of data-intensive computing. Overcoming the challenges of data-intensive computing has required optimization of data movement across multiple levels of memory hierarchies, and these considerations have become even more important as we prepare for exascale computing. The approaches taken to address these challenges include (a) fast data output from a large simulation for future processing/archiving; (b) minimization of data movement across caches and other levels of the memory hierarchy; (c) optimization of communication across nodes using fast and low-latency networks, and communication optimization; and (d) effective co-design, usage and optimization of system components from architectures to software.” [1]

Over the past decade, a new paradigm for scientific discovery is emerging due to the availability of exponentially increasing volumes of data from large instruments such as telescopes, colliders, and light sources, as well as the proliferation of sensors and high-

throughput analysis devices. Further, data sources, analysis devices, and simulations are connected with current-generation networks that are faster and capable of moving significantly larger volumes of data than in previous generations. These trends are popularly referred to as big data. However, generation of data by itself is of not much value unless the data can also lead to knowledge and actionable insights. Thus, the fourth paradigm, which seeks to exploit information buried in massive datasets to drive scientific discovery, has emerged as an essential complement to the three existing paradigms. The complexity and challenge of the fourth paradigm arises from the increasing velocity, heterogeneity, and volume of data generation." [1]

Jyväskylän yliopiston IT-tiedekunta on kehittänyt data-analyysin koulutusta ja tutkimusta systemaattisesti yhteistyössä matematiikan ja tilastotieteen laitoksen sekä kansainvälisten huippuyliopistojen, Tel Avivin ja Yalen yliopistojen, kanssa. IT-tiedekunnan monialainen osaaminen luo hyvän pohjan alan koulutuksen ja tutkimuksen kehittämiseen. Kehitystyötä tullaan jatkamaan sekä metodien että sovellusten osalta (käsittäen kyberturvallisuuden, hyperspektrikameran kuvantamisdatan käsittelyn, big datan sekä Business Intelligent:iin liittyvän datan käsittelyn).

3.2 Big datan tutkimus

3.2.1 Perusteita

Suurien datamassojen tutkimusta toteutetaan tilastotieteessä, laskennallisten tieteiden ja sovelletun matematiikan alueilla.

Tilastotieteen tutkimusaloja ovat mm:

- Spatiaalinen tilastotiede tarkastelee paikkatietoaineistojen tilastollista analysointia ja mallinnusta sekä tilastollista kuva-analyysia
- Aikasarja-analyysin tutkimus kohdistuu tila-avaruusmallien ja monimuuttujaisten aikasarjamallien teoriaan ja metodikehitykseen.
- Rakenneyhtälömallien tutkimus on kompleksisten monimuuttujaisten aineistojen ja pitkittäisaineistojen mallinnusta
- Parametrittomien ja robustien monimuuttujamenetelmien tutkimus on merkki- ja järjestyslukuvektoreihin perustuvien monimuuttujamenetelmien teoreettista kehitystyötä
- Biometrian ja ympäristötilastotiede on tutkimusalue, joka sisältää populaation mallinnusta ja vesistöjen ekologisen tilan arviointia

Erityisen kiinnostava tutkimusalue ovat spatiaaliset mallit. Paikkatietoon perustuvia ennustemalleja voidaan tuottaa päätöksentekijöitä varten muodostamalla datasta jakaumia, kasautumia, riippuvuuksia ja poikkeamia. Havainnoista voidaan tehdä päätelmiä luoda hypoteeseja jatkoanalyysiin. Mobiiliteknologian alueella paikkatiedolla on yhä suurempia sovellusalueita. Spatiaalisten mallien rakentamisen tavoitteena on

tutkittavan ilmiön ymmärtäminen, jotta voidaan rakentaa malli ilmiön käyttäytymisen ennustamista varten.

Laskennallisten tieteiden tutkimusaloja ovat matemaattinen mallintaminen, luotettava malli- ja datapohjainen simulointi, optimointi, adaptiiviset ja tehokkaat numeeriset laskentamenetelmät, epävarmuuden huomioiminen numeerisessa simuloinnissa, hajautettujen systeemien säätö, spline ja spline wavelet tekniikat signaalin ja kuvankäsittelyssä, dynaamiset systeemit ja nanoelektroniikan mallinnus.

Sovelletun matematiikan tutkimusaloja ovat mm. diskreetti matematiikka, matemaattinen mallintaminen, funktionaalianalyysi, mitta- ja integraaliteoria ja kompleksianalyysi.

Jyväskylän yliopistossa data-analyysin tutkimusaloja ovat analysointimenetelmien kehittäminen, erityisesti numeriikka ja massiivisen datan luokittelutekniikat, hyperspektrikameran datan analysointitekniikoiden kehittäminen ja tekniikan soveltaminen sen osa-alueilla, kuten solubiologia, lääketiede, ympäristötiede, maa- ja metsätalous, kemialliset aseet, rikospaikkatutkimustekniikka. Lisäksi tutkimukseen liittyviä yhteistyöhankkeita on mm. fysiikan ja aivotutkimuksen alueilla.

3.2.2 Data-analyysi Jyväskylän yliopiston kesäkoulussa 2013

Jyväskylän kansainvälisessä kesäkoulussa elokuussa 2013 professori Amir Averbuch piti kurssin (2 op) aiheesta: *Advanced Methods for Classification of Big High Dimensional Data*.

3.2.3 Big data-analyysiin liittyviä hankkeita

Data-analyysin alueella on toteutunut ja käynnissä seuraavia hankekokonaisuuksia.

Data-analyysin vuonna 2013 päättynyt hanke:

- Professori Timo Hämäläinen: Suurien moniulotteisten datajoukkojen järjestäminen ja analysointi, HIDE-hanke (1.1.2012–31.12.2013, Tekes)

Data-analyysin käynnissä olevia tutkimushankkeita:

- Professori Pekka Neittaanmäki: New System for Cyber Attacks Protection of Critical Infrastructures 2012–2014, CAP-projekti (1.11.2012–31.10.2014, Tekes)

- Professori Jari Veijalainen: Tiedonkaivuu sosiaalisesta mediasta, MineSocMed-projekti (1.9.2013–31.8.2017, SA) tarkoitus kehittää sosiaalisen median analyysialgoritmeja.

- Professori Timo Hämäläinen: Kiinteistöautomaatiojärjestelmien datan älykäs analysointi, KIIAUDATA-hanke (1.1.2013–31.12.2014, Tekes)

- Professor Amir Averbuch: MeBUD: Methods for Big Unstructured High Dimensional Data (1.8.2014 - 30.7.2018, haettu rahoitusta SA)

3.3 Big data-analyysin opetus

Suurien tietomassojen analyysin opiskelu toteutetaan kolmen maisteriohjelman sisällä, joissa opiskelija voi profiloitua data-analyysiin. Tietotekniikan laitoksella toteutetaan laskennallisten tieteiden ja sovelletun matematiikan maisteriohjelmat ja Matematiikan ja tilastotieteen laitoksella toteutetaan tilastotieteen maisteriohjelma.

3.3.1 Tilastotieteen maisteritutkinto

Tilastotieteen maisteriopinnot sisältävät sekä teoreettisia opintoja että tilastotieteen sovelluksia ja tähtäävät ammattitilastotieteilijän taitoon. Tilastotiede kehittää malleja ja menetelmiä numeerisen havaintoaineiston keräämiseen, kuvaamiseen ja analysointiin ja tähän liittyvään laskennalliseen toteuttamiseen. Tilastotieteellä on kiinteä yhteys lähes kaikkiin empiiristä tutkimusta tekeviin tieteenaloihin: tilastollisia menetelmiä sovelletaan niin informaatioteknologiassa, bio- ja ympäristötieteissä, taloustieteessä, lääketieteessä kuin yhteiskunta- ja kasvatustieteissäkin. Tilastotieteessä on kysymys reaalimaailman ilmiöiden mallintamisesta ja sen osaamista tarvitaan yhä enemmän yhteiskunnassa ja elinkeinoelämässä, missä tutkimusaineistojen ja tietovarantojenanalyyseilla ja mallinnuksella halutaan tuottaa jalostettua tietoa päätöksenteon tueksi.

Tilastotieteen opetuksen tavoitteena on antaa valmiudet edustavien havaintoaineistojen keräämiseen, aineistojen kuvaamiseen ja analysointiin sekä yleensä numeerisesti mitattavissa olevienilmiöiden pätevään tilastolliseen mallintamiseen. Tilastotieteellä on käytettävissä erilaisia analysointityökaluja data-analyysin toteuttamiseen.

3.3.2 Sovelletun matematiikan maisteritutkinto

Sovelletun matematiikan avulla pyritään ratkaisemaan tosielämän ongelmia. Sovelletun matematiikan tavoitteena on mallintaa erilaisia ilmiöitä, kuvailla niitä ja yrittää ymmärtää niitä. Sovelletun matematiikan opiskelussa yhdistyy tieteellisen laskennan käsitteet ja menetelmät, joita käytetään kysymyksiin, jotka ilmentyvät matematiikan ja muiden tieteenalojen rajapinnoissa. Jyväskylän yliopistossa opinnoissa keskitytään sellaisiin osa-alueisiin, kuten funktionaalianalyysi, mitta- ja integraaliteoria, kompleksianalyysi, numeerinen analyysi, optimointi ja simulointi.

Valmistunut maisteri hallitsee laaja-alaisesti sovelletun matematiikan ja tieteellisen laskennan käsitteitä ja menetelmiä, joita käytetään itsenäisen ajattelun ja tutkimuksen perustana. Ymmärtää matematiikan ja lähitieteenalojen rajapintojen tietoihin liittyviä laskennallisia kysymyksiä ja tarkastelee niitä ja uutta tietoa kriittisesti.

Sovellettu matematiikka tuottaa matemaattisia työkaluja data-analyysin toteuttamiseen.

3.3.3 Laskennallisten tieteiden maisteritutkinto

Laskennallisten tieteiden maisterikoulutuksessa käsitellään laaja-alaisesti tilastotieteen, numeerisen laskennan ja ohjelmoinnin käsitteitä ja menetelmiä. Laskennallisten tieteiden maisteri tuntee jatkuvan ja diskreetin simuloinnin periaatteet ja sovelluskohteet. Hän osaa listata jatkuvien simulointimallien tavallisimmat diskretisointimenetelmät ja niiden tehokkaan toteuttamisen perusperiaatteet moderneissa tietokonearkkitehtuureissa. Lisäksi hän osaa nimetä yksi- ja monitavoitteisen epälineaarisen optimoinnin periaatteet ja ratkaisumenetelmät.

Hän kykenee muodostamaan tekniikan ja luonnontieteiden ilmiöille matemaattisia simulointimalleja sekä osaa rakentaa mallien ratkaisemiseen tehokkaat ohjelmistot aliohjelmakirjastoja tai vastaavia valmiita komponentteja hyödyntäen. Hän osaa muodostaa ja ratkaista numeerisesti simulointimalleihin pohjautuvia optimointitehtäviä.

Laskennalliset tieteet antavat erilaisia numeerisia työkaluja data-analyysin toteuttamiseen.

3.3.4 Data-analyysin osaamisprofiili

Edellä kuvatut kolme maisterikoulutusta antavat opiskelijalle mahdollisuuden profiloitua suurten datamassojen analyysiin kunkin tieteenalan näkökulmasta ja tutkimustyökaluja hyväksikäyttäen.

Data-analyysissä opetetaan ja tutkitaan menetelmiä ja lähestymistapoja, joilla eritavoin kerätystä tiedosta (data) pyritään muodostamaan malleja ja korkeampaa tai tarkempaa informaatiota. Opetuksessa korostuu keskeisinä tekijöinä datan kerääminen, käsittely ja visualisointi.

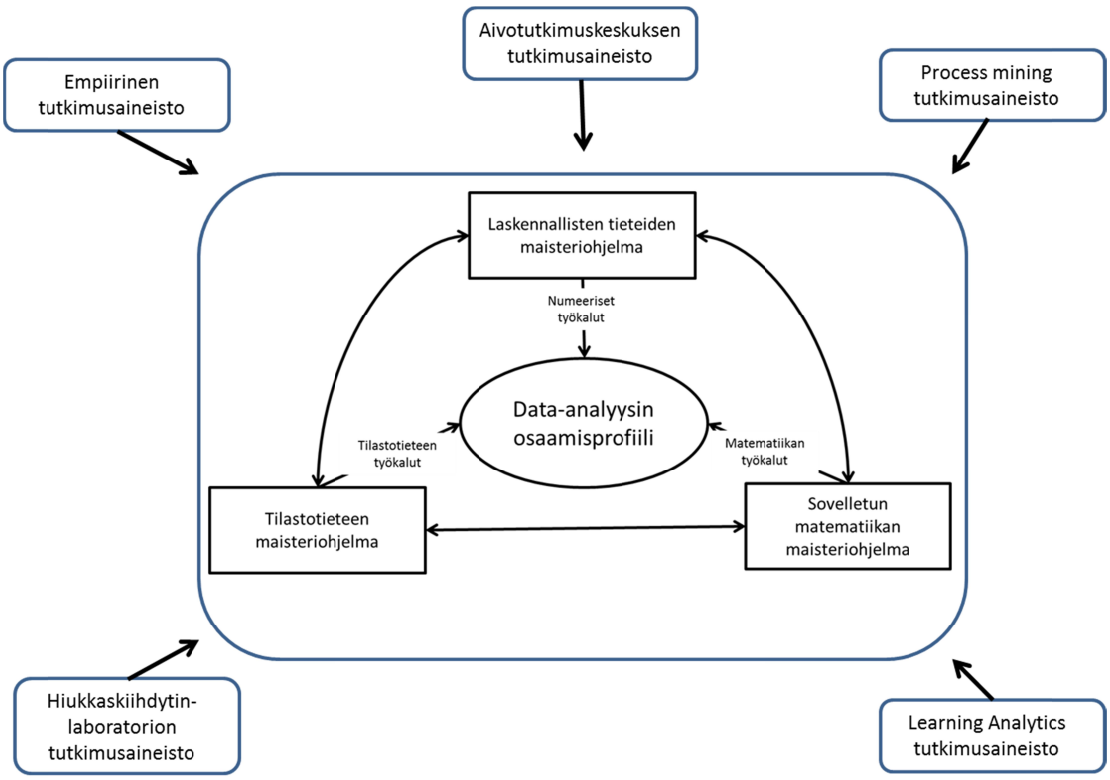
Data-analyysiin erikoistuva opiskelija rakentaa osaamisprofiilinsa mukaisesti opintosuunnitelmansa eri kurssikokonaisuuksista, kuten;

- tilastotieteen kursseja
- sovelletun matematiikan kursseja
- tietoliikennetekniikan kursseja
- sensoriverkkoihin liittyviä kursseja
- kokonaisarkkitehtuurin rakentamiseen liittyviä kursseja

Data-analyysin maisterikoulutus vastaa muuttuvan maailman tilanteeseen, jossa suurien data-aineistojen automaattisesta analysoinnista on tullut keskeinen työkalu useilla aloilla. Koulutuksen tavoitteena on antaa opiskelijoille data-analyysiin liittyvää erikoisosaamista sekä tilastollisista menetelmistä että niiden soveltamisesta tietokoneympäristöön.
Jyväskylän yliopiston laaja-alainen ja monitieteellinen toimintaympäristö antaa mahdollisuuden data-analyysin opiskelijoille käyttää hyväkseen erilaisia datamassoja, kuten:

- oppimiseen liittyvä data (Learning Analytics)
- hiukkaskiihdyttimen tuottama havaintoaineisto
- aivotutkimusyksikön tuottama havaintoaineisto
- erilaiset prosessien tuottamat data-aineistot (Process mining)
- muu empiirinen tutkimusaineisto

Kuvassa 2 on esitetty data-analyysin osaamisprofiilin rakentuminen Jyväskylän yliopiston toimintaympäristössä.



KUVA 2 Data-analyysin osaamisprofiilin rakentuminen

Lukuvuonna 2014–2015 tarjotaan seuraavia erityisesti data-analyysiin liittyviä kursseja:

- ITKST47 Advanced Anomaly Detection: Theory, Algorithms and Applications, 5 op
- ITKST48 Advanced Persistence Threat, 5 op, Advanced Persistence Threat exploitation cycle
- ITKA204 Tietokannat ja tiedonhallinnan perusteet, 4 op

- TIES445 Tiedonlouhinta, 5 op
- TJTSM61 Business Analytics and Big Data Management, 5 op
- ITKXX Big Data Engineering

3.3.5 Big data-analyysin jatkokoulutus

Vuosina 2013–2014 julkaistiin seuraavat data-analyysiä käsittelevät väitöskirjat:

- Guy Wolf: Big high-dimensional data analysis with diffusion maps.
- Ilkka Pölönen: Discovering knowledge in various applications with a novel hyperspectral imager
- Tuomo Sipola: Knowledge discovery using diffusion maps
- Limor Gavish: Memcached - noSQL and big data databased particularly for caching
- Mikhail Zolotukhinin: On data mining applications in mobile networking and network security
- Antti Juvonen: Intrusion Detection Applications Using Knowledge Discovery and data Mining
- Gil Shabat: Computationally Efficient Tools for Big Data Processing
- Moshe Salhov: Manifold learning from structured kernels and out of sample extensions
- Hannu-Heikki Puupponen: Unmixing Methods in Novel Applications of Spectral Imaging

LIITE 1 Kyberturvallisuuden kurssitarjonta 2014–2015

1. Organisaatiolähtöinen osaamisprofiili:

Pakolliset syventävät opinnot
- ITKST41 Kybermaailma ja turvallisuus, 5 op
- ITKST53 Ohjelmistoturvallisuus, 3-5 op
- TIES327 Tietoverkkoturvallisuus, 3-5 op
- ITKS452 Requirements Engineering, 5 op

Valinnaiset syventävät opinnot

Cyber Security Technology
- ITKST42 Anomaly Detection, 5 op
- ITKST47 Advanced Anomaly Detection: Theory, Algorithms and Applications, 5 op
- ITKST 48 Advanced Persistence Threat, 5 op

Systems Security
- ITKST 50 System vulnerabilities, 5 op
- ITKST 51 Operating system security 1, 5 op
- ITKST 52 Operating system security 2, 5 op

Cyber Conflict
- ITKST44 Kybermaailma ja kansainvälinen oikeus, 5 op
- ITKST45 Introduction to cyber conflict, 5 op
- ITKST46 Cyber defence strategy analysis, 5 op
- ITKST 55 Kyberhyökkäys ja sen torjunta, 5 op

2. Organisaatiolähtöinen osaamisprofiili:

Pakolliset syventävät opinnot
- ITKST41 Kybermaailma ja turvallisuus, 5 op
- ITKST43 Information Security Management, 5 op
- ITKS452 Requirements Engineering, 5 op
- TIES327 Tietoverkkoturvallisuus, 3-5 op

Valinnaiset syventävät opinnot
- ITKST40 Yhteiskunta ja informaatioturvallisuus, 5 op
- ITKSTXX Advanced Course on Information Security Management, 5 op
- ITKSTXX Information privacy, 5 op

- ITKSTXX Secure Systems Design, 5 op
- TJTSXX Research Methods 5 op
- ITKST45 Introduction to cyber conflict, 5 op
- ITKST46 Cyber defence strategy analysis, 5 op

LIITE 2 Kyberturvallisuuden jatko-opiskelijoita

Nimi	Lai-tos	Aihe
Alasuutari Minna	TKTL	Pedagogisesti hyvin suunnitellun tietoturvakoulutuksen vaikutus tietoturvakäyttäytymiseen
Aronsson Wilhelmiina	TKTL	Kyberturvallisuuden johtaminen - häiriötilanteiden hallinta
Bilozerov Oleksandr	TTL	Organizational Information Security in SMES
Gavish, Limor	TTL	Caching in Web Based Applications
Ghanbari, Hadi	TKTL	Theoretical Approaches to Information Systems Development Process
Jiang, Hemin	TKTL	To Explain and Affect Cyberloafing Behavior: a Control Balance Perspective
Jin, Xueyu	TKTL	Continuous usage of micro-blogging websites: a territoriality perspective
Juvonen, Antti	TTL	Tietoturvan parantaminen käyttäen tiedonlouhintaa ja koneoppimista
Kiperberg, Michael	TTL	The Benefits of Virtualization in Security and Performance
Kokkonen, Tero	TTL	Cyber Security Situation Picture
Krawczyk, Piotr	TKTL	Tietohallinnon riskit ja niiden hallinta
Kronqvist Jyrki	TTL	Ulkoistuksen tietoturva
Kuem, Jungwon	TKTL	The Value of Constructive Non Work related Web Use (NWWU) on Individual Performance
Miettinen, Matti	TKTL	Kyberturvallisuuden näkökulmasta yrityksenvaikutuskeskeisen ja palvelusuuntaisen yhteistoimintakyvyn kognitiotieteellinen tutkimus
Nabi, Syed Irfan	TKTL	Grounded Ontology - A Text Coding Approach to Ontology Development for Human Behavior Aspects of Information Security
Nykänen, Riku	TTL	Small and medium enterprise information security strategies
Rajamäki, Jyri	TTL	Designing future Public Protection and Disaster Relief (PPDR) vehicles ICT integration: the communications layer
Resh, Amit	TTL	Enforcing Trust in Modern Environments
Rotbart, Aviv	TTL	Anomaly Detection and Classification of High-Dimensional Patterns via Localized Substructures
Salhov, Moshe	TTL	Non-Scalar Spectral Embedding for Data Analysis and Dictionary Based Out-of-Sample Extension
Shabat, Gil	TTL	Developing and Applying Low Rank Methods for Big Data Manipulation
Tambe Ebot, Alain Claude	TKTL	Fall Prey to Spear Phishing Attacks: A Process Approach to Theory Development

Vahdani Amoli, Payam	TTL	Intelligent Anomaly Detection Systems
Wartiainen, Pekka	TTL	Visuaalisia analyysimenetelmiä tietomurtohyökkäysten louhintaan
Woods, Naomi	TKTL	Examing memory and IS security awareness to improve security performance

LÄHTEET

[1] Synergistic Challenges in Data-Intensive Science and Exascale Computing, Summary Report of the Advanced Scienti Computing Advisory Committee (ASCAC) Subcommittee, March 2013, http://science.energy.gov/~/media/40749FD92B58438594256267425C4AD1.ashx)

[2] Työ- ja elinkeinoministeriö, 21 polkua Kitkattomaan Suomeen, ICT 2015 -työryhmän raportti 17.1.2013, http://www.tem.fi/ajankohtaista/julkaisut/julkaisujen_haku/21_polkua_kitkattomaan_suomeen.98249.xhtml

[3] Valtioneuvoston periaatepäätös, Suomen kyberturvallisuusstrategia, 24.1.2013, www.yhteiskunnanturvallisuus.fi.

Informaatioteknologian tiedekunnan julkaisuja
No. 15/2014

ISBN 978-951-39-6042-1 (verkkoj.)
ISSN 2323-5004

JYVÄSKYLÄN YLIOPISTO