2014/2500M-0014

ISSN 1797-3457 (verkkojulkaisu)
ISBN (PDF) 978-951-25-2639-0

TIIVISTELMÄRAPORTTI

Virheenkorjauskoodien tunnistus signaalitiedustelussa

Prof. Patric Östergård, TkT Jussi Poikonen, Ville Kuvaja
Yhteystiedot: Patric Östergård, Tietoliikenne- ja tietoverkkotekniikan laitos, Aalto-yliopiston sähkötekniikan korkeakoulu, PL 13000, 00076 Aalto, Suomi • Email: Patric.Ostergard@aalto.fi

Tutkimus keskittyy digitaalisissa tiedonsiirtojärjestelmissä käytettävien virheenkorjauskoodien tunnistamiseen tuntemattomista signaaleista. Aihe liittyy MATINEn vuoden 2014 tutkimuksen painopisteeseen *Elektronista- ja informaatiosodankäyntiä tukevat teknologiat* ja erityisesti *Tiedustelu- suojautumis-, häirintä- ja valvontajärjestelmien ratkaisuihin.* Tutkimuksen päätavoitteet ovat virheenkorjauskoodien tunnistamisen rajoitusten ja vaatimusten teoreettinen analyysi ja eri koodien tietoturvallisuuden tarkastelu, koodien tunnistamismenetelmien kehitys sekä virheenkorjauskoodien tunnistamismenetelmien käytännön toteutus, testaus ja demonstrointi. Projektin tulokset ovat tarpeellisia nykyaikaisten digitaalisten tietoliikennejärjestelmien signaalitiedustelussa ja hyödyksi myös esimerkiksi tietoturvallisten viestintäjärjestelmien suunnittelussa.

1 Johdanto

Nykyaikainen nopea digitaalinen tiedonsiirto perustuu olennaisesti tehokkaaseen virheenkorjauskoodaukseen. Digitaalitekniikan kehitys on mahdollistanut jo matkaviestinlaitteissakin käsiteltävän niin monimutkaisia virheenkorjauskoodeja, että virheettömän viestinnän nopeudet lähestyvät tiedonsiirtokanavien teoreettisia rajoja. Signaalitiedustelun kannalta koodien monimutkaistuminen tarkoittaa, että mikäli viestintäjärjestelmässä käytetty virheenkorjauskoodi ja sen parametrit eivät ole tiedossa, on lähetetyn tiedon purkaminen vaikeaa; käytännössä vaikka vastaanotetun viestin koodausmenetelmä olisi tiedossa, kaikkien mahdollisten koodivaihtoehtojen läpikäyminen on yleensä mahdotonta. Toisaalta esimerkiksi ohjelmistoradiotekniikka mahdollistaa virheenkorjauskoodauksen konfiguroimisen dynaamisesti, lähettimen ja vastaanottimen ohjelmakoodia muokkaamalla. Seuraavassa kuvaillaan tarkemmin digitaalisten tietoliikennejärjestelmien signaalitiedustelun haasteita ja erityisesti hankkeen aiheena olevaa tuntemattomien virheenkorjauskoodien parametrien selvittämistä. Hankkeessa tehty tutkimus liittyy MATINEn vuoden 2014 tutkimuksen painopisteeseen Elektronista- ja informaatiosodankäyntiä tukevat teknologiat, ja tämän sisällä erityisesti tiedustelu-, suojautumis-, häirintä- ja valvontajärjestelmien ratkaisuihin.

Signaalitiedustelun tavoitteena on selvittää vastaanotetuista radioviesteistä niin paljon tietoa kuin mahdollista. Yksinkertaisten tunnistettavien suureiden, kuten lähetyksen kantotaajuuden ja kaistanleveyden, jälkeen on digitaalisessa järjestelmässä selvitettävä suuri määrä lähetyksen parametreja. Käytännössä ensin on löydettävä vastaanotetun viestin modulaatio eli järjestelmässä hyödynnettävien aaltomuotojen joukko. Tämän jälkeen on tunnistettava virheenkorjauskoodi ja siihen liittyvät parametrit. Lähetettävän tiedon lomittaminen on virheenkorjaukseen liittyvä, mutta siitä erillinen operaatio, jossa tiettyyn koodisanaan sisältyvät bitit hajautetaan ajallisesti eri osiin radiolähetystä pitkien virhepurskeiden ehkäisemiseksi. Jotta virheenkorjauskoodi voidaan purkaa, on ensin selvitettävä lomituksessa käytetty menetelmä. Käytännössä on usein tarpeellista tunnistaa sa-

manaikaisesti lomittelumenetelmä ja virheenkorjauskoodi, mikä on nykyaikaisilla koodeilla yleisesti vaikea tehtävä.

Signaalitiedustelun matemaattista teoriaa ei ole toistaiseksi olemassa. Kaikki tiedonsiirtomenetelmät kuitenkin vaikuttavat lähetettävien signaalien tilastollisiin ominaisuuksiin näkyen erilaisina korrelaatiorakenteina, joita signaalitiedustelija saattaa käyttää hyväkseen. Hankkeessa tarkasteltiin näitä tilastollisia rakenteita ja niiden hyödyntämistä virheenkorjauskoodien ja niihin liittyvien parametrien tunnistamisessa. Lähtökohtaisesti tiedustelun onnistumista rajaa signaali-kohinasuhde ja mahdollisesti signaalin lyhyt kesto, joka johtaa vähäiseen näytteiden lukumäärään. Tiedustelussa yritetään sokeasti estimoida suuri määrä kuvauksia, jolloin lähtökohtaisesti voi olla olemassa tapauksia, joissa hyvälaatuistakin signaalia joudutaan vastaanottamaan pitkään luotettavien tulosten saamiseksi. Myös tarvittava laskentakapasiteetti kasvaa virheenkorjauksen monimutkaisuuden funktiona.

2 Tutkimuksen tavoite ja suunnitelma

2.1 Työn tavoite

Työssä keskityttiin yksinomaan virheenkorjauskoodien ja niihin liittyvän lomituksen tunnistamiseen, joten tutkimuksen ulkopuolelle rajattiin esimerkiksi hajaspektrikoodien, modulaatiomenetelmien sekä salauskoodien tunnistus ja käsittely (Kuva 1). Nämä aiheet ovat olemassa olevassa kirjallisuudessa laajemmin käsiteltyjä kuin tässä tutkimuksessa tarkasteltava ongelma. Käytännön toteutusten kannalta olennaiset signaalimodulaation ja tiedonsiirtokanavan vaikutukset otettiin huomioon olettamalla, että vastaanotetut signaalit voivat sisältää tiedonsiirtokanavan ja epäideaalisen demodulaation aiheuttamia virheitä.

Kuva 1. Tiedonsiirtojärjestelmän yksinkertaistettu rakenne. Tässä siirtokanavaan oletetaan sisältyvän mm. signaalin modulointi, mahdolliset hajaspektrimenetelmät sekä radiosignaalin eteneminen. Tässä tutkimuksessa tarkasteltiin kanavadekoodauksen toteuttamista ilman esitietoja käytetystä koodauksesta. Yllä punaisella merkityt siirtokanava ja kryptografisen salauksen purkaminen rajattiin työn ulkopuolelle.

Edellä mainitun rajauksen perusteella työssä oletettiin, että käsiteltävät signaalit ovat muodoltaan digitaalisia, toisin sanoen diskreettejä ja äärellisestä joukosta arvojaan saavia. Erityisesti käsiteltiin binäärisignaaleja, eli ykkösten ja nollien sekvenssejä, jotka edustavat virheenkorjauskoodattua dataa. Datasta käytössä olevasta ennakkotiedosta tehtiin tarvittaessa tutkimusta helpottavia olettamuksia, esimerkiksi että koodisanojen pituudet ovat tunnettuja tai että käytössä olevat koodit tai lomitusmenetelmät kuuluvat johonkin ennestään tiedossa olevaan kandidaattijoukkoon.

Tutkimukselle asetettiin seuraavat kolme päätavoitetta:

1. Virheenkorjauskoodien tunnistamisen rajoitusten ja vaatimusten teoreettinen analyysi
   - Informaatioteoreettisessa mielessä ydinkysymys on, paljonko eri koodeilla ja tiedonsiirtokanavilla saadaan informaatiota käytössä olevasta koodista vastaanotettua bittiä kohden.
   - Kompleksisuusteorian kannalta olennaista on, miten monimutkainen laskennallinen ongelma tietyn koodin selvittäminen on. Tähän liittyvänä käytännön tu-

loksena voidaan arvioida eri koodityyppien tietoturvallisuutta.

2. Koodien tunnistamismenetelmien kehitys
   - Määritellään yllä mainittujen teoreettisten rajoitusten puitteissa joukko testikoodeja, joita pyritään tunnistamaan eri kanavaolosuhteissa. Lähtökohtaisesti oletetaan käytössä olevan suhteellisen hyvälaatuinen siirtokanava, mutta kohinan ja binäärimuotoisten virheiden vaikutukset tarkasteltujen menetelmien suorituskykyyn otetaan huomioon.
3. Tunnistamismenetelmien testaus ja demonstrointi
   - Tunnistamismenetelmiä toteutetaan Matlab-ohjelmistolla ja testataan simuloimalla satunnaisesti generoituja koodeja ja lähetettävää dataa.

2.2 Toteutussuunnitelma

Tutkimus aloitettiin aiheen kattavalla kirjallisuustutkimuksella, ongelmakentän määrittelyllä sekä analyysillä siitä, miten laskennallisesti kompleksista eri virheenkorjauskoodityyppien tunnistus on sekä paljonko mittadataa voidaan teoriassa odottaa tarvittavan eri koodityyppien luotettavaan tunnistamiseen. Näiden tarkastelujen perusteella valittiin virheenkorjauskoodit, joiden tunnistamista tarkasteltiin lähemmin. Erityisesti työssä keskityttiin Low Density Parity Check (LDPC) -koodeihin, jotka ovat merkittäviä nykyaikaisissa tietoliikennejärjestelmissä.

Hankkeen vanhempi tutkimushenkilökunta, professori Patric Östergård ja tohtori Jussi Poikonen, keskittyivät työssä erityisesti yllä mainittuun teoreettiseen analyysiin sekä tunnistusmenetelmien ja niiden testauksen suunnitteluun. Eri menetelmien ja koodien toteuttamisen ja käytännön testaamisen tueksi määriteltiin diplomityö, jonka tekijä, Ville Kuvaja, osallistui hankkeen tulosten toteuttamiseen yhteistyössä vanhempien tutkijoiden kanssa. Hankkeessa tarkasteltujen koodintunnistusmenetelmien toteutuksia testattiin tietokonesimulaatioiden avulla.

3 Aineisto ja menetelmät

Tutkimuksen teoreettisen viitekehyksen muodostavat tilastotiede, todennäköisyyslaskenta sekä informaatioteoria. Tutkimuksen tärkeimmät metodit ovat teoreettinen analyysi, numeerinen analyysi sekä kokeelliset menetelmät. Tutkimusaiheen teoreettisessa analyysissa sovellettiin erityisesti todennäköisyyslaskentaa sekä informaatioteoriaa esimerkiksi koodien tunnistamisen perustavanlaatuisten rajoitusten sekä vaadittavan laskennan kompleksisuuden arviointiin. Numeerisessa analyysissa hyödynnettiin tilastollisia menetelmiä eri virheenkorjauskoodien tuottaman datan analysointiin, luokitteluun ja tunnistamiseen.

Tärkeä osa työtä oli myös aiheen tutkimuksen nykytilan selvittäminen ja tähän liittyvä kirjallisuustarkastelu. Materiaalit tähän tarkasteluun hankittiin pääasiassa IEEE:n julkaisutietokannasta. Tutkimuksen tuloksia testattiin kokeellisin menetelmin, erityisesti soveltamalla koodien tunnistusmenetelmiä simuloituun dataan. Simulaatioissa tunnistettavat koodit generoitiin satunnaisesti.

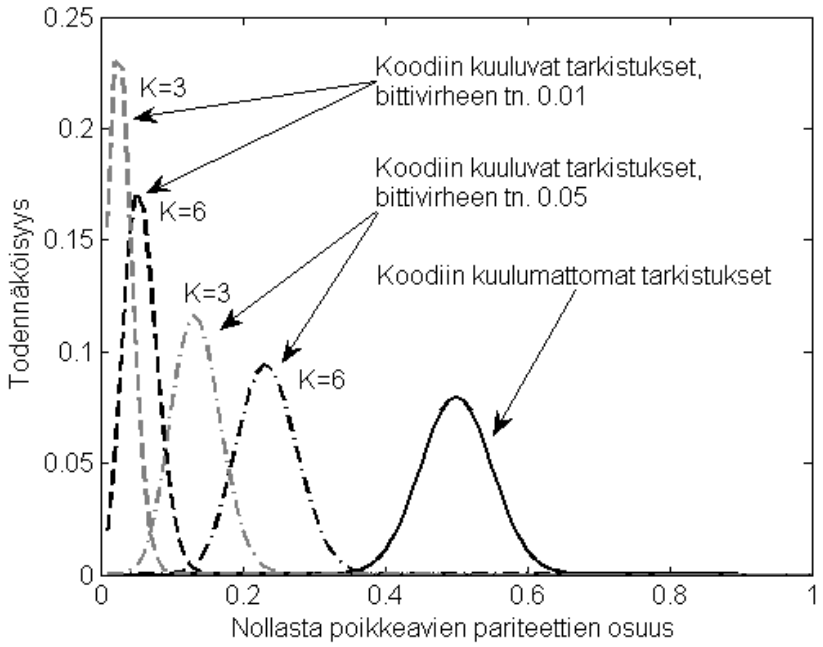
4 Tulokset ja pohdinta

Tutkimuksen tärkeimmät tulokset ovat tieto tutkimusaiheen kansainvälisestä nykytilasta, analyysit erityisesti LDPC-koodirakenteiden säännöllisyyksien hyödyntämisestä koodien tunnistamisongelman helpottamiseksi sekä tunnistusalgoritmien toteutukset ja testaus

simuloidulla datalla.

Tutkimuksen alussa tehdystä kirjallisuustutkimuksesta saatiin kattava kuva aiheen kansainvälisestä tämänhetkisestä tuntemuksesta. Tästä yhteenvetona voi todeta, että perinteisten lohko- ja konvoluutiokoodien, näiden yhteydessä suoritettavan lomituksen sekä klassisista koodeista johdettujen modernien turbo- ja LDPC-koodien tunnistamiseen on olemassa toimivia menetelmiä, mutta näiden menetelmien toimivuus ja tehokkuus riippuvat voimakkaasti tunnistettavan koodin ja lomitusmenetelmän ominaisuuksista sekä virheiden todennäköisyydestä vastaanotetussa datassa. Teorettisesti voidaan arvioida esimerkiksi kuinka pitkään ja millä signaalin voimakkuudella tunnistettavaa dataa on pystyttävä vastaanottamaan, jotta koodin tunnistaminen on mahdollista tietyllä todennäköisyydellä.

Käytännössä esimerkiksi LDPC-koodeja tunnistettaessa on pystyttävä tilastollisesti erottelemaan koodiin kuuluvien ja kuulumattomien pariteettitarkistusten tulosten jakautumia, jotka lähestyvät toisiaan vastaanotetun datan virhetodennäköisyyden ja koodin painokertoimen (kuhunkin pariteettitarkistukseen osallistuvien bittien lukumäärän) kasvaessa. Kuva 2 havainnollistaa tätä ilmiötä; käytännössä koodin painokertoimen tai vastaanotetun datan virhetodennäköisyyden kasvaessa tarvitaan siis enemmän vastaanotettua dataa, jotta etsittyyn koodiin kuuluvat pariteettitarkistukset pystytään tunnistamaan luotettavasti.

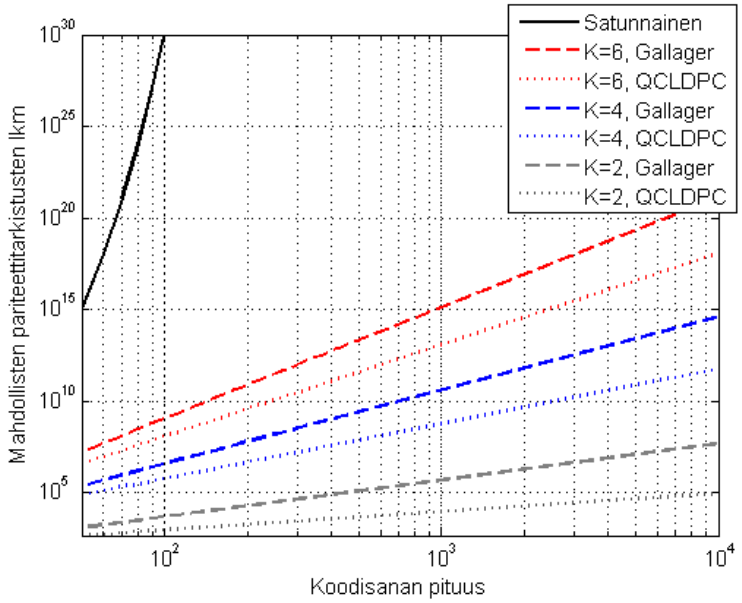


Kuva 2. Esimerkkejä tiedonsiirtokanavasta aiheutuvien bittivirheiden vaikutuksesta LDPC-koodien tunnistamiseen. Koodin tunnistamiseksi on erotettava toisistaan tunnistettavaan koodiin kuuluvien ja kuulumattomien pariteettitarkistusten tulosten jakautumat. Nämä jakautumat lähestyvät toisiaan virhetodennäköisyyden ja koodin painokertoimen K kasvaessa, jolloin niiden erottamiseksi toisistaan tietyllä varmuudella tarvitaan enemmän vastaanotettua dataa.

Työssä analysoitiin tarkemmin LDPC-koodien tunnistamista sekä mahdollisuuksia tehostaa koodien tunnistamista erilaisten koodien rakenteellisten säännönmukaisuuksien perusteella. Todettiin esimerkiksi, että mikäli LDPC-koodin pariteettitarkistusmatriisi on kvasisyklinen, säännöllinen sarakepainojensa suhteen tai systemaattinen, on mahdollista merkittävästi helpottaa koodin tunnistamista, mikäli tunnistusalgoritmissa otetaan huomioon nämä piirteet. Lisäksi työssä toteutettiin ja raportoitiin simulaatioita LDPC-koodien

tunnistusalgoritmin toiminnasta.

Esimerkkinä LDPC-koodirakenteen vaikutuksesta koodin tunnistamisen laskennalliseen kompleksisuuteen kuvassa 3 on esitetty mahdollisten eri pariteettitarkistusten lukumäärän suuruusluokka täysin satunnaiselle binäärikoodille, Gallager-tyyppisille LDPC-koodeille sekä kvasisyklisille LDPC-koodeille (QCLDPC), joiden pariteettitarkistusmatriisi koostuu syklisesti siirretyistä identiteettimatriiseista. Kuvasta voi todeta, että tunnistamisessa läpikäytävien eri pariteettitarkistusten joukko kasvaa nopeasti koodin pituuden ja painokertoimen funktiona, mutta koodin säännöllinen rakenne pienentää läpikäytävää joukkoa merkittävästi.



Kuva 3. Pahimman tapauksen laskennallisen kompleksisuuden suuruusluokan arvioita Gallager-tyyppisten LDPC-koodien, kvasisyklisten QCLDPC-koodien sekä täysin satunnaisesti valittujen koodien tunnistamisessa. Muuttuja K on koodin painokerroin, ts. kuinka monta bittiä kuhunkin pariteettitarkistukseen osallistuu. Täysin satunnaisen koodin tapauksessa K olisi noin puolet koodisanan pituudesta.

Hankkeessa tehdyllä tutkimuksella on hyödyntämismahdollisuuksia sotilaallisessa maanpuolustuksessa, erityisesti signaalitiedustelussa sekä tietoturvallisten viestintäjärjestelmien suunnittelussa. Työssä tarkastellut virheenkorjauskoodit ovat merkittävä osa nykyaikaisten tietoliikennejärjestelmien toimintaa, ja työn tulosten perusteella voidaan todeta, että tyypillisiä moderneissa järjestelmissä käytettäviä virheenkorjauskoodeja on mahdollista tunnistaa tarkasteltujen menetelmien avulla. Tällaiset menetelmät ovat käytännössä välttämättömiä signaalitiedustelussa, jonka tavoitteena on rakenteeltaan tuntemattomassa digitaalisessa tiedonsiirtojärjestelmässä lähetettävän tiedon selvittäminen. Tutkimuksen käytännön soveltuvuuden testaamiseksi työtä olisi kuitenkin laajennettava soveltamalla tunnistusmenetelmiä käytännön järjestelmiin ja mitattuun dataan.

5 Loppupäätelmät

Työn tulokset toteuttavat hankkeelle asetetut päätavoitteet. Työssä selvitettiin nykytiedon valossa virheenkorjauskoodien tunnistamisen rajoituksia ja vaatimuksia. Työn perusteella on mahdollista toteuttaa käytännössä tyypillisten tuntemattomien virheenkorjauskoodattujen signaalien tunnistamista sekä toisaalta määritellä virheenkorjauskoodirakenteita ja niihin liittyviä parametreja siten, että koodien tunnistaminen on vaikeaa. On kui-

tenkin huomattava, että vaikeasti tunnistettaviksi suunnitellut koodit tyypillisesti vaikeuttavat myös viestivän järjestelmän toimintaa esimerkiksi kasvattamalla signaalinkäsittelyn viivettä, vastaanottimessa tarvittavan muistin kokoa sekä dekoodauksen laskennallista kompleksisuutta.

Yhtenä työn alkuperäisenä tavoitteena oli käsitellä tarkemmin sekä LDPC- että turbokoodien tunnistamista. Tämä tarkempi käsittely rajattiin lopulta kuitenkin LDPC-koodeihin; turbokoodien tunnistusalgoritmien syvällisempi tarkastelu on yksi mahdollinen jatkoaihe tutkimukselle. Tavoitteena oli myös, että työn tuloksista laadittaisiin diplomityö sekä kansainvälinen konferenssi- tai lehtiartikkeli. Aiheesta tehty diplomityö on parhaillaan viimeistelyvaiheessa, ja tutkimuksen tuloksina saatiin konferenssiartikkeliin sopivaa aineistoa. Mahdollinen julkaisukanava tuloksille olisi esimerkiksi IEEE MILCOM, johon jätettävien artikkelien seuraava määräaika on huhtikuussa 2015.

Selkeitä jatkoaiheita tutkimukselle ovat esimerkiksi tunnistusalgoritmien toteutukset muille kuin tässä työssä tarkastelluille virheenkorjauskoodeille, koodien rakenteiden säännöllisyyksien hyödyntämisen jatkotutkimus sekä toteutukset käytännön järjestelmissä ja käyttöolosuhteissa tallennetulla datalla. Mahdollinen jatkotutkimus lienee kuitenkin järkevää integroida osaksi laajempaa signaalitiedustelutyötä, ottaen huomioon käytännössä tunnistettavien signaalien ja järjestelmien todennäköiset ominaisuudet.

6 Tutkimuksen tuottamat tieteelliset julkaisut ja muut mahdolliset raportit

Tutkimuksesta valmistuu diplomityö, joka julkaistaan alkuvuodesta 2015:

Ville Kuvaja, *Identification of Error Correction Codes in Signals Intelligence.* Diplomityö, Aalto-yliopisto, Sähkötekniikan korkeakoulu, 2015.

Tämä työ sisältää hankkeessa tehdyn kirjallisuustutkimuksen sekä kuvaukset testattujen tunnistamismenetelmien toteutuksesta ja simulaatiotuloksista.