2017/2500M-0067

ISSN 1797-3457 (verkkojulkaisu)
ISBN (PDF) 978-951-663-010-9

TIIVISTELMÄRAPORTTI

Algebralliset menetelmät virheenkorjauskoodin tunnistuksessa

Dosentti Jyrki Lahtonen, Turun yliopisto, matematiikan ja tilastotieteen laitos,
lahtonen@utu.fi

1. Johdanto

Radioteitse tapahtuvassa digitaalisessa kommunikointiketjussa on useita vaiheita. Biteiksi koodattu viesti on yleensä kryptattu. Sen jälkeen kryptattu viesti suojataan kanavan aiheuttamilta virheiltä virheenkorjauskoodilla (FEC = forward error correction). Tällä tavalla tuotettu bittivirta on vielä muunnettava radioaalloiksi valitsemalla tarkoituksenmukainen joukko käytettäviä aaltomuotoja, ja sopimalla bittikombinaation ja radioaallon välinen vastaavuus.

Kun radiokanavassa sitten havaitaan tällainen viesti, sen analysoija pyrkii purkamaan kyseisen ketjun käänteisessä järjestyksessä. Tehtävä voidaan karkeasti jakaa seuraaviin vaiheisiin: 1) käytetyn modulaatiomenetelmän ja aaltojoukon tunnistaminen, 2) aaltomuotojen bittileimojen määrääminen, 3) käytetyn virheenkorjauskoodin tunnistaminen, 4) FEC-koodin syötteen määrääminen, 5) kryptauksen purkaminen.

Tällainen analysointitehtävä tulee vastaan kognitiivisessa radiossa, johon osallistuva laite kuuntelee ympärillä olevaa radiospektriä, ja pyrkii tunnistamaan sieltä eri standardien mukaisia radiojärjestelmiä. Samaan ongelmaan törmätään myös sotilaallisessa signaalitiedustelussa. Näissä kahdessa tilanteessa on myös vivahde-eroja muun muassa sen suhteen kuinka paljon aikaa ja laskentaresursseja analysointiin voidaan käyttää, ja millaista ennakkotietoa analysoitavan signaalin rakenteesta analysoijalla on olemassa.

Turun yliopiston matematiikan laitokselle on kertynyt FEC-koodien algebralliseen rakenteeseen liittyvää kompetenssia, ja yhdessä sidosryhmiemme (PVTIEDL) kanssa pystyimme isoloimaan käytännön signaalitiedustelun kannalta relevantteja ongelmia, jotka tulevat vastaan kuvatun ketjun vaiheessa 3. PVTIEDL olikin rahoittanut aikaisemmin kahta opinnäytetyötä, joissa perehdyttiin menetelmään tunnistaa konvoluutiokoodi, sen jokin yksinkertainen lomittelija tai punkteerauskuvio. Konvoluutiokoodit valikoituivat kohteeksi A) niiden runsaan käytön ja B) sopivan rikkaan algebrallisen rakenteen vuoksi.

Lomittelemattoman konvoluutiokoodatun viestin voi tunnistaa etsimällä sen duaalikoodin pieneen ikkunaan rajoittuvia sanoja. Tästä suoriudutaan joko arvaamalla tällainen sana, ja testaamalla arvauksen oikeellisuus. Hieman algebrallisempi ja tehokkaampi tapa tuottaa haettuja duaalikoodin sanoja on selvittää näytesanan ja sen viivästettyjen versioiden väliset lineaariset riippuvuudet. Jälkimmäinen tapa tosin edellyttää, ettei näytteessä ole bittivirheitä. Tilastollisena menetelmänä edellinen sietää jonkin verran huonolaatuisemman näytteen – voidaan jopa käyttää ns. pehmeitä bittejä. Joka tapauksessa yksinkertaisen taulukkolomittelijan käyttö voidaan tässä ottaa huomioon. Se tuottaa tietenkin kertoimen testattavien vaihtoehtojen määrälle.

Konvoluutiokoodia käytettäessä voidaan mukautua muuttuviin kanavaolosuhteisiin sopimalla käyttäjien kesken tiettyjä punkteerauskuvioita. Tällöin radiokelin ollessa hyvä ja virheenkorjaustarpeen vastaavasti matalampi viestintä tehostuu. Tällainen punkteerattu koodi säilyttää edelleen punkteerattoman ”äitikoodin” algebrallisen rakenteen pääasialliset piirteet. Näiden ollen viestin tunnistaminen konvoluutiokoodilla tuotetuksi onnistuu

kuten edellä. Punkteerauskuvio vaikuttaa konvoluutiokoodin rakenteen yksityiskohtiin ennustettavalla tavalla, jota voidaan hyödyntää testeissä. Jälkimmäisessä mainituista opinnäytetöistä perehdyttiin tämän tehtävän yksityiskohtiin, ja tuotettiin algoritmi, joka tulostaa lyhyehkön listan vaihtoehtoja. Lisäksi tehtiin havainto, että oikea vaihtoehto tuotti aina parametreiltaan yksinkertaisimman koodin. Tätä heuristiikkaa testattiin STANAG- ja MilSpec-standardeissa spesifioituihin konvoluutiokoodeihin ja niiden punkteerauskuvioihin. Testatuissa tilanteissa heuristiikka toimi.

2. Tutkimuksen tavoite ja suunnitelma

Toteutetussa hankkeessa haluttiin jatkaa johdannossa kuvailluissa opinnäytetöissä aloitettua työtä. Tavoitteeksi asetettiin lisätä repertuaariimme toinen usein käytetty FEC-koodien perhe, Reedin-Solomonin koodit. RS-koodeja käytettäessä viestit on ensin koodattu äärellisen kunnan alkioiksi, ja vasta sen jälkeen ne muutetaan biteiksi. Tästä syystä detektioketjun 3. vaiheessa on otettava huomioon seuraavat seikat:

- Käytetyn kunnan koko ei lähtökohtaisesti ole analysoijan tiedossa. Yleisimmin on käytössä (esimerkiksi CD-levyjen, digi-TV:n ja QR-koodien kanssa) 8-bittinen kunta, mutta muihinkin vaihtoehtoihin tulee varautua. 4-bittistä kuntaa voidaan käyttää lyhyissä koodilohkoissa esimerkiksi kommunikaatioprotokollien kättelyvaiheessa signaloimaan radiokehyksen rakennetta kuvaavia parametreja tms.
- Sen jälkeen, kun käytetyn kunnan koko tunnetaan, on edelleen olemassa useita mielekkäitä tapoja muuntaa kunnan alkio bittikombinaatioksi. Tavoitteeksi asetettiin mahdollisuuksien mukaan rajata näitä vaihtoehtoja ja/tai jopa ratkaista kysymys näytteen tai näytteiden perusteella.

Konvoluutiokoodien osalta asetettiin tavoitteiksi:

- Selittää johdannossa mainituissa opinnäytetöissä havaitut anomaliat. Muutamissa tehdyissä testiajoissa ja simulaatioissa oli saatu hieman yllättäviä tuloksia, joiden syytä ei heti keksitty.
- Täydentää konvoluutiokoodin havaitseva ohjelmistomme kattamaan myös tailbiting-tekniikalla tuotetut muunnelmat, ja analysoimaan mahdollisia testeissä esiintyviä ongelmatilanteita.
- Uudenlaisena problematiikkana päätettiin yrittää analysoida mahdollisuutta yhdistää detektioketjun vaiheita 2 ja 3. Koska vaiheessa 2 joudutaan tekemään arvaus bittikuvioiden ja aaltomuotojen väliseksi vastaavuudeksi, analysoijalla on tarvetta työkalulle, joka kertoo jotakin tehdyn arvauksen oikeellisuudesta. Esimerkkinä tällaisesta ongelmasta käytin seuraavaa. Radioteitse kommunikoidessa käytetään usein QAM (= quadrature amplitude modulation) -aakkostoa. QAM-aakkosten I- ja Q-haaroja vastaavat bittikombinaatiot määritellään yleensä Gray-koodin avulla. Kuitenkin Gray-koodeja on useita, joten analysoija joutuu arvaamaan käytetyn Gray-koodin. Tavoitteena oli analysoida väärän arvauksen seurauksia, ja miettiä voiko analysoija tehdä diagnoosin väärästä arvauksesta.

Oli ennakoitavissa, että näihin tehtäviin valittavia opiskelijoita pitää täsmäkouluttaa. Tältä osin varauduttiin siihen, että allekirjoittaneen opetusohjelmaan lisättiin (virkaan kuuluvana työnä) konvoluutiokoodien kurssi keväälle 2017. Reedin-Solomonin koodien osaltakin vastaavaa kurssitustarvetta saattaisi ilmetä, mutta oli todennäköistä, että tarvittavan äärellisten kuntien algebran tuntevia opiskelijoita löytyisi. Täsmäkoulutettujen opiskelijoiden joukosta olisi kuitenkin valittavissa sopivia henkilöitä. Selvää kuitenkin oli jo suunnitteluvaiheessa, että hankkeen varsinaiseen työhön voidaan ryhtyä vasta loppukeväästä.

Projektin toteutusvaiheessa noudatettaisiin aiheen julkaisuihin tutustuminen jälkeen perinteistä sykliä: implementointi → testiajot → tulosten analysointi → algoritmien parantaminen/uusien lähestymistapojen etsiminen. Haastavampiin tehtäviin (useampiin syklin kierroksiin vaaditaan luovaa ajattelua) valittaville henkilöille tuli varata tarpeellinen määrä aikaa.

3. Aineisto ja menetelmät

Kuten johdannossa mainituissa opinnäytetöissä käytimme testiaineistona NATOn standardoimia konvoluutiokoodeja ja niistä tuotettuja tailbiting-koodeja. Tarvittavat detektiorutiinit ja testipenkit koodattiin Mathematica-ympäristössä, jolloin pääsimme hyödyntämään aiempien hankkeiden tuotoksia. RS-koodien osalta testimateriaali tuotettiin joko itse tai valmiita MATLAB-ympäristön rutiineja käyttäen.

RS-koodeja koskevassa osuudessa tutkijani (FM Toni Hotanen ja FM Taneli Lehtilä) aloittivat implementoimalla kirjallisuudesta löytyneitä algoritmeja, ja analysoivat testiajojen tuloksia algebran tuntemuksensa avulla. Hankkeen loppuvaiheessa Lehtilä teki laajempia testiajoja RS-koodien kvasisyklisen rakenteen tarkaksi selvittämiseksi etsien viitteitä siitä, miten kunnan määrittelevä polynomi paljastuisi.

Konvoluutiokoodien puolella lyhyehkössä osaprojektissa tutkijani (FM Antti Peltola ja FM Teemu Pirttimäki) laajensivat aiemmat detektioalgoritmit tunnistamaan myös tailbiting-koodit. Tarkempi raportti ohessa. Pidemmässä osaprojektissa FM Anni Hakanen syventyi kuvattuun ongelmaan analysointiketjun vaiheiden 2 ja 3 yhdistämisestä. Hänen menetelmänään oli testata, mitä aiemmat analysointityökalumme kertoisivat näytteestä, jossa bitit on tuotettu väärin arvatulla menetelmällä.

4. Tulokset ja pohdinta

Projektin tuloksina:

1. Testattiin aiemmin tuotettu konvoluutiokoodien detektion tekevä koodi perusteellisemmin. Havaitut anomaliat saivat selityksensä ja ne korjattiin. Hankkeen kannalta tämän vaiheen merkitys oli siinä, että samalla opiskelijat tuli perehdytettyä problematiikkaan.
2. Tutkittiin Reedin-Solomonin koodin detektion problematiikkaa:
    a. Implementoitiin ja testattiin algoritmi RS-lohkon löytämiseksi kuntasymboleiksi käännetystä syötteestä. Lisäksi algoritmia laajennettiin siten, että se selviää tunnistustehtävästä myös silloin kun signaalinäyte muodostuu muutamasta taulukkolomitellusta RS-lohkosta. Kun meillä on oikea arvaus tavasta muuntaa bittikombinaatio aakkostokunnan alkioksi, niin osoittautui, että tunnistus on käytännössä täysin luotettava. Edelleen RS-koodin lohkon löytäminen näytteen keskeltä oli suhteellisen suoraviivainen ratkaistava aakkostokunnan diskreettiä Fourier-muunnosta hyödyntäen. Tässä koodin parametrien selvittämiseen riitti melkein aina yhden lohkon mittainen näyte, jonka toinen näyte varmisti erittäin luotettavaksi. Pieni epävarmuus aiheutui siitä, että RS-koodin sana on noin puolen prosentin todennäköisyydellä myös hieman pienemmän RS-koodin sana.
    b. Implementoitiin ja testattiin algoritmi RS-lohkon löytämiseksi binäärisestä syötteestä sekä implementoitiin algoritmi, joka etsii kunnan määrittelevän polynomin kokeilemalla vaihtoehdot. FM Toni Hotanen toteutti tämän MATLABilla. Hänen algoritminsa vaati syötteeksi 8 lohkon mittaisen näytteen. Tällä edelly-

tyksellä lohkojen sykliset siirrot virittävät koko koodin mahdollistaen sen parametrien identifioinnin.

c. Implementoitiin ja testattiin vaihtoehtoinen algoritmi kohdan 2b ongelmaan, joka yrittää polynomialgebran avulla päätellä määrittelevän polynomin suoraan näytteestä. Analysoitiin jäljelle väistämättä jäävää rakenteen monikäsitteisyyttä (FM Taneli Lehtilän raportti ohessa). Samasta syystä kuin b-kohdassa tässä tarvitaan 8 lohkon mittainen näyte.

3. Kehitettiin, implementoitiin ja testattiin eräs menetelmä tailbiting konvoluutiokoodin lohkojen tunnistamiseksi (FM Teemu Pirttimäen raportti ohessa).
4. Perehdyttiin muutamiin ongelmatilanteisiin, joihin törmätään, kun yhdistetään detektioketjun vaiheet 2 ja 3 konvoluutiokoodin osalta:

   a. Analysoitiin mahdollisuuksia tunnistaa ja korjata väärä arvaus QAM-moduloidun signaalin aaltomuotojen leimauksessa käytetystä Gray-koodista. Hieman yllättäen osoittautui, että väärä arvaus Gray-koodista ei estä koodin tunnistamista konvoluutiokoodiksi vanhoilla työkaluilla. Tämä tulos pystyttiin myös selittämään konvoluutiokoodien algebran avulla. Kehiteltiin myös heuristiikka oikean Gray-koodin löytämiseksi. Tyypillisesti väärä Gray-kuvaus johti tilanteeseen, jossa muodostuva konvoluutiokoodi oli joko epäilyttävän monimutkainen tai sitten ominaisuuksiltaan kelvoton, ja näin analysoijan eliminoitavissa (FM Anni Hakasen raportti ohessa).

   b. Analysoitiin vastaava ongelma käytettäessä differentiaalista QAM-modulaatiota. Tähänkin kehitettiin samantapainen heuristiikka (FM Anni Hakasen raportti ohessa).

Olen itse erittäin tyytyväinen yo. luettelon kohtiin 2c, 4a ja 4b, ja tyytyväinen kohtiin 1, 2a, 2b ja 3. Minkään kohdan oikeellisuudesta ei minulla ole epäilyksiä. Osoittautui, että kohdan 3 tehtävä voidaan ratkaista myös toisella tavalla, joka on laskennallisesti tehokkaampi, mutta siinäkin tapauksessa tuloksemme täydentää toista lähestymistapaa.

Kehitettyjen algoritmien hyödynnettävyydestä käytännön signaalitiedustelussa en valitettavasti pysty sanomaan paljoa. Kohdissa 3 ja 4 käytimme testiaineistoina STANAG- ja MilSpec-standardeissa määriteltyjä FEC-koodeja. On mahdollista, että jokin muu sotilaallinen toimija käyttää niitä tai niiden muunnelmia, jolloin algoritmimme purevat. Tiedossani ei ole, onko tällaisia koodilohkoja havaittu signaalitiedustelussa.

Tulosten tieteellisestä merkittävyydestä arvioni on projektin kesto huomioon ottaen erittäin myönteinen. Kohdan 2c tulokset sopisivat hyvin esiteltäviksi alan huippukonferenssissa, ja pienen lisäjalostuksen jälkeen myös artikkelina. Kohtiin 4a ja 4b tehty analyysi ja sen ratkaisemiseksi kehitetty heuristiikka johtavat mielestäni kansainvälisen tason julkaisuun. Kummankin tuloksen kohdalla julkaisuksi työstäminen vaatii tekijältä vielä lisäpanostusta.

5. Loppupäätelmät

Mielestäni hankkeen tavoitteisiin päästiin hyvin. Edellisen kohdan huipputulokset jopa osin ylittivät omat odotukseni ongelmatilanteiden analyysin osalta.

Oma hankejohtamiseni ei onnistunut aivan yhtä hyvin. Kohtaa 3 työstettiin valtaosin loma-aikanani, ja tulokset olisivat saattaneet olla parempia, jos olisin ehtinyt osallistua ratkaisumenetelmien valintaan.

Meillä kattavana oletuksena on ollut näytesignaalin virheettömyys – oletimme ettei yksikään bitti ollut tuhoutunut kanavavirheen seurauksena. Sotilaallisessa signaalitieduste-

lussa voi useinkin materiaalia olla tarjolla riittävästi tällaisen korkealaatuisen näytteen löytämiseksi. Olisi kuitenkin myös tarjolla mielenkiintoisia mahdollisuuksia laajentaa menetelmämme kattamaan myös tilanteita, jossa näytteessä on kohtuullinen määrä kanavavirheitä.

Edellisen kappaleen johtopäätös tarjoaa ilmeisen jatkohankemahdollisuuden. On kuitenkin todettava, että toisen yhtä pätevän henkilöstön kouluttaminen siihen työhön ei tapahdu nopeasti. Sekä FM Hakasen että FM Lehtilän käytettävyys on tätä kirjoitettaessa auki, joten mahdollinen jatkohanke siirtyy myöhemmäksi.