Tampereen yliopisto

Maria Koskio

LAAJENEVAT GRAAFIT JA HASHFUNKTIOT

Informaatioteknologian ja viestinnän tiedekunta
Pro gradu -tutkielma
Matematiikka
Joulukuu 2020

TIIVISTELMÄ

Maria Koskio: Laajenevat graafit ja hashfunktiot
Pro gradu -tutkielma
Tampereen yliopisto
Matematiikka
Joulukuu 2020

Tutkielmassa tarkastellaan laajenevia graafeja ja hashfunktioita. Aluksi käsitellään yleisesti hashfunktioita ja niiden tietoturvaominaisuuksia sekä graafi- ja spektraaligraafiteorian perusteita. Täten annetaan tarvittavat pohjatiedot varsinaisten aiheiden ymmärtämiseen. Tämän jälkeen siirrytään laajenevien graafien perheiden pariin, jotka määritellään kahdella eri tavalla; ensin laajenemissuhteiden perusteella ja sitten spektrien aukkojen kokoihin perustuen. Jälkimmäisen osoitetaan olevan ekvivalentti edeltävän kanssa. Lopuksi käsitellään laajenevia hashfunktioita keskittyen niiden tietoturvaominaisuuksiin sekä kokonaiskuvan hahmottamiseen. Yleisen tarkastelun jälkeen esitellään laajenevat ZT-, LPS- ja Pizerin hashfunktiot.

Avainsanat: hashfunktio, graafiteoria, laajenevien graafien perhe, laajeneva graafi, laajeneva hashfunktio, ZT-hashfunktio, LPS-hashfunktio, Pizerin hashfunktio

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

Sisältö

1 Johdanto **4**

2 Hashfunktioiden perusteet **6**
2.1 Hashfunktiot ja niiden tietoturvaominaisuudet 6
2.2 SHA-perheen hashfunktiot 7
2.3 Tietoturvaominaisuuksien riippuvuuksia 10
2.4 Väsytyshyökkäyksiä hashfunktioita vastaan 13

3 Graafiteorian perusteet **16**
3.1 Graafit, solmut ja särmät 16
3.2 Solmun ja graafin aste sekä säännöllisyys 18
3.3 Spektraaligraafiteorian perusteita 20
3.4 Erilaisia polkuja graafissa 23
3.5 Graafien ominaisuuksia 26

4 Laajenevien graafien perheet **28**
4.1 Laajenemissuhde ja sen merkitys 28
4.2 Spektrin aukon koon merkitys 30

5 Laajenevia hashfunktioita **35**
5.1 Laajenevat hashfunktiot ja niiden tietoturvaominaisuudet 35
5.2 Esimerkkejä laajenevista hashfunktioista 37

1 Johdanto

Päivittäin käytetään useita salasanoja, mutta kuinka moni on koskaan pysähtynyt miettimään niiden tallentamisen tietoturva-aspekteja. Knospen [9] mukaan monet käyttöjärjestelmät ja sovellukset eivät tallenna salasanoja sellaisenaan, vaan niiden hasharvon. Tämä hasharvo saadaan antamalla salasana jollekin toivottavasti tietoturvalliselle hashfunktiolle, joka sitten muuttaa sen lyhyemmäksi hasharvoksi, joka ei paljasta itse salasanaa. Kun käyttäjä haluaa kirjautua sisään, verrataan hänen kirjoittamansa salasanan hasharvoa tallennettuun hasharvoon. Myös monet muut nykyään yhä yleisemmät toiminnot, kuten digitaaliset allekirjoitukset, viestien autentikointi ja kryptovaluutat, hyödyntävät hashfunktioita. Kryptografiaa ja hashfunktioita eivät tarvitsekaan enää pelkästään sotilaat ja vakoojat, vaan aivan jokainen, joka viestii internetin välityksellä.

Tässä tutkielmassa käsitellään erityisesti laajenevien graafien perheisiin perustuvia hashfunktioita eli laajenevia hashfunktioita. Aihe on ajankohtainen, sillä viimeisten kahdenkymmenen vuoden aikana ovat Priyadarsinin [21] mukaan tutkijat osoittaneet erityistä kiinnostusta hyödyntää kryptografiassa nimenomaan graafiteoriaa, mistä monet NP-kovat ongelmat nousevat. Myös laajenevien hashfunktioiden tietoturva vastaa tiettyjä graafiteoreettisia ongelmia sekä laajenevien graafien ominaisuuksia. Painopiste onkin laajenevien graafien ja hashfunktioiden ymmärtämisessä sekä tietoturvassa sen sijaan, että paneuduttaisiin niiden suorituskykyvaatimuksiin ja tehokkuuteen.

Luvussa 2 käsitellään hashfunktioita yleisesti antaen samalla pohjatietoja ymmärtää laajenevia hashfunktioita. Se jakautuu neljään alalukuun. Niistä ensimmäisessä määritellään hashfunktiot ja niiden tietoturvaominaisuudet. Toisessa perehdytään laajalti käytettyihin SHA-perheen hashfunktioihin esimerkkinä tavanomaisista hashfunktioista. Tietoturvaominaisuuksien riippuvuuksia toisistaan sekä matemaattisista ongelmista tutkitaan tarkemmin kolmannessa alaluvussa ja neljännessä alaluvussa tietoturvaominaisuuksia koetellaan erilaisin väsytyshyökkäyksin. Yleisesti hashfunktioihin perehtynyt lukija voikin halutessaan aloittaa suoraan seuraavasta luvusta.

Näkökulma vaihtuu kokonaan, kun siirrytään lukuun 3. Se jakautuu viiteen alalukuun, joissa valmistaudutaan laajenevien graafien perheiden tarkasteluun perehtymällä graafiteorian perusteisiin. Ensimmäisessä alaluvussa annetaan määritelmät ja esitystavat graafeille, solmuille ja särmille, jotta päästään alkuun. Toisessa alaluvussa tarkastellaan solmun ja graafin astetta sekä säännöllisyyttä ja luodaan siten tarpeellisia termejä ja työvälineitä jatkon kannalta. Lineaarialgebran perustiedot voivat helpottaa lukijaa, kun kolmannessa alaluvussa syvennytään spektraaligraafiteorian perusteisiin, jotka ovat erityisen tärkeitä laajenevien graafien tarkastelun kannalta. Tärkeitä ovat myös neljännessä alaluvussa käsiteltävät erilaiset polut, jotka mahdollistavat kulkemisen graafeissa. Viidennessä alaluvussa määritellään graafeille vielä muutamia lisäominaisuuksia, kuten yhtenäisyys ja isomorfisuus. Yleisesti graafi- ja spektraaligraafiteoriaan perehtynyt lukija voi puolestaan halutessaan jättää tämän

luvun välistä.

Luvussa 4 jatketaan graafien parissa keskittyen laajenevien graafien perheisiin ja luodaan siten loputkin valmiudet ymmärtää laajenevia hashfunktioita. Tämä luku jakautuu kahteen alalukuun, joista ensimmäinen aloitetaan muutamalla määritelmällä, kuten raja ja laajenemissuhde, joiden avulla sitten määritellään laajenevien graafien perheet. Toisessa alaluvussa hyödynnetään spektraaligraafiteoriaa ja annetaan vaihtoehtoinen määritelmä laajenevien graafien perheelle spektrien aukkojen kokoihin perustuen. Määritelmät myös osoitetaan keskenään ekvivalenteiksi. Lisäksi määritellään Ramanujanin graafien perheet, jotka riippuvat spektrien aukkojen ko'oista. Laajenevista graafeista riittäisi kerrottavaa enemmänkin, mutta tässä tutkielmassa keskitytään laajenevien hashfunktioiden kannalta oleellisiin asioihin.

Vihdoin luvussa 5 yhdistetään hashfunktiot sekä laajenevat graafit ja siten päästään tutustumaan laajeneviin hashfunktioihin. Niitä käsitellään kahdessa alaluvussa ja tavoitteena on antaa idea sekä mahdollistaa kokonaiskuvan hahmottaminen. Ensimmäisessä alaluvussa aloitetaan satunnaispoluista, jotka johdattavat laajenevien hashfunktioiden ja niiden tietoturvaominaisuuksien pariin. Tietoturvaominaisuuksia tarkasteltaessa huomioidaan niin hashfunktion kuin laajenevan graafinkin ominaisuuksien vaikutukset niihin. Toiseen alalukuun on valittu laajenevat ZT-, LPS- ja Pizerin hashfunktiot esimerkeiksi ja lisäksi tarjotaan lähdeviitteet viiteen muuhun. Valitut esimerkit tarjoavat hyvän otannan erilaisista laajenevista hashfunktioista, sillä ne ovat kaikki erilaisia tietoturvaominaisuuksiltaan ja perustuvat eri laajenevien graafien perheisiin. Siksi esimerkkien ymmärtämistä voi helpottaa matriisilaskennnan, diskreetin matematiikan ja elliptisten käyrien perusteiden tunteminen ennestään.

2 Hashfunktioiden perusteet

Tässä luvussa tarkastellaan yleisesti kryptografisia hashfunktioita (cryptographic hash functions), joihin jatkossa viitataan hashfunktioina. Stinsonin mukaan [24] hashfunktioita käytetään luomaan lyhyitä ikään kuin sormenjälkiä pidemmille viesteille. Näin jos viestiä muutetaan, niin sen sormenjälki ei enää täsmää. Tämä mahdollistaa sen, että voidaan varmistua viestin muuttumattomuudesta, vaikka se olisi tallennettuna turvattomaan paikkaan. Hashfunktioita käytetäänkin kryptografiassa paitsi huomaamaan muutoksia tallennetuissa viesteissä (assurance of data integrity) niin myös digitaalisissa allekirjoitusmenetelmissä (digital signature schemes) ja viestien autentikoinnissa (message authentication codes).

2.1 Hashfunktiot ja niiden tietoturvaominaisuudet

Määritellään ensin viesti ja bitti. Näiden avulla määritellään yleisesti (avaimeton) hashfunktio, jota Stinsonin [24] mukaan yleensä käytetään huomaaman muutoksia tallennetuissa viesteissä sekä digitaalisissa allekirjoitusmenetelmissä.

Määritelmä 2.1. Viesti on äärellinen merkkijono, joka voidaan esittää unicode-standardin avulla heksadesimaalilukuna ja siten edelleen binäärilukuna.

Määritelmä 2.2. Bitti on binääriluvun yksittäinen numero, jonka arvo on aina joko 0 tai 1. Binääriluku, jonka pituus on n bittiä, on n bittinen. Kaikki n bittiset luvut muodostavat joukon $\mathbb{B}_n = \{0, 1\}^n$.

Määritelmä 2.3. Vrt. [22, luku 18.1] ja [24, s. 120]. Olkoon M joukko mahdollisia viestejä ja n positiivinen kokonaisluku. Kun $h\colon M \to \mathbb{B}_n$ on helposti laskettavissa kaikilla $m \in M$, niin se on hashfunktio. Joukon $\mathbb{B}_n$ alkiot ovat niitä vastaavien viestien hasharvoja.

Hashfunktion helposti laskettavuus on vahvasti tilanteesta riippuvainen ja siksi sitä ei ole määritelty tässäkään tarkasti. Arkielämässä se saattaa tarkoittaa, että tavanomainen tietokone selviytyy laskutoimituksesta kymmenesosasekunneissa. Toisaalta esimerkiksi arkaluontoista tiedustelutietoa käsiteltäessä voidaan hyväksyä joidenkin minuuttien laskenta-aika suurelta palvelinsalilta.

Seuraavaksi määritellään salausavain ja avaimellinen hashfunktio, jolla voidaan laajentaa mahdollisten hasharvojen joukkoa tai muuttaa viestien jakautumista samassa joukossa. Erityisesti avaimelliset hashfunktiot ovat Stinsonin [24] mukaan käytössä viestien autentikoinnissa.

Määritelmä 2.4. Salausavain on tiettyyn funktioon liittyvä viestistä riippumaton muuttuja, joka voi olla joko salainen tai julkinen.

Määritelmä 2.5. Vrt. [24, s. 120]. Olkoon $h\colon M \to \mathbb{B}_n$ hashfunktio. Kun k on hashfunktioon liittyvä salausavain ja $h_k\colon M \to \mathbb{B}_n$ on helposti laskettavissa kaikilla $m \in M$, niin h_k on avaimellinen hashfunktio.

Esimerkki 2.1. Olkoon M joukko, joka sisältää kaikki suomen kielen sanat. Määritellään $h\colon M \to \mathbb{B}_1$, jolla kaikilla $m \in M$ pätee $h(m) = 1$. Tällöin h on selvästi kaikin tavoin helposti laskettavissa, joten se on hashfunktio. Nähdään myös, että $h(''\text{kana}'') = 1$, joten sanan ”kana” hasharvo on 1. Kuitenkin koska kaikkien sanojen hasharvo on 1, niin kyseinen hashfunktio ei ole millään tavalla tietoturvallinen eikä edes hyödyllinen.

Esimerkin pohjalta huomataan, että aiemmin esitettyjen määritelmien lisäksi tarvitaan hashfunktioille joitakin tietoturvaominaisuuksia, jotta niistä on käytännössä hyötyä aiemmin mainituissa käyttötarkoituksissa. Yleisimmät tietoturvaominaisuudet määritellään seuraavaksi ja ne ovat samat sekä avaimettomille että avaimellisille hashfunktioille.

Määritelmä 2.6. Vrt. [22, luku 18.1] ja [24, s. 121]. Olkoon $h\colon M \to \mathbb{B}_n$ hashfunktio. Tällöin sen tietoturvaominaisuuksiin viitataan seuraavin termein.

1. **Alkukuvaresistenssi** (preimage resistance): Olkoon $b \in \mathbb{B}_n$. Jos on vaikeaa löytää $m \in M$, jolla $h(m) = b$, niin hashfunktio h on alkukuvaresistentti.

2. **Toisen alkukuvan suhteen resistenssi** (second preimage resistant): Olkoon $m \in M$. Jos on vaikeaa löytää $m' \in M$, jolla $m \neq m'$ ja $h(m) = h(m')$, niin hashfunktio h on toisen alkukuvan suhteen resistentti.

3. **Yksisuuntaisuus** (one-way): Jos hashfunktio h on sekä alkukuvan että toisen alkukuvan resistentti, niin se on yksisuuntainen.

4. **Törmäysresistenssi** (collision resistance): Jos on vaikeaa löytää $m, m' \in M$, joilla $m \neq m'$ ja $h(m) = h(m')$, niin hashfunktio h on törmäysresistentti.

Tietoturvaominaisuuksien vaikeasti löydettävyys on käytännössä helposti laskettavuuden vastakohta ja siksi sekin vaihtelee tilanteen mukaan. Ei kuitenkaan voida vaatia, että kyseiset ongelmat olisivat mahdottomia ratkaista, sillä äärellisten joukkojen ollessa kyseessä aina voidaan kokeilla kaikki mahdolliset ratkaisut. Tämän tyyppisiä hyökkäyksiä hashfunktioiden tietoturvaominaisuuksia vastaan kutsutaan väsytyshyökkäyksiksi ja niistä yleisimmät määritellään alaluvussa 2.4. Tietoturvaominaisuuksien välisiä riippuuvuuksia puolestaan tarkastellaan alaluvussa 2.3.

2.2 SHA-perheen hashfunktiot

Seuraavassa esimerkissä määritellään SHA-0 niminen hashfunktio, joka on ollut käytössä vuosina 1993-1995 ja mahdollisesti hieman sen jälkeenkin. Tämän jälkeen kerrotaan muistakin SHA-perheen hashfunktioista ja niiden kehityksestä.

Esimerkki 2.2. Vrt. [1, s. 56-58]. SHA-0 on seuraavanlainen hashfunktio mahdollisten viestien joukolta $\mathcal{M}$ joukolle $\mathbb{B}_{160}$. Olkoon $M \in \mathcal{M}$ binäärilukuesitys viestistä, jonka hasharvo, $SHA(M)$, halutaan laskea. Apumuuttujien ja vakioiden arvoja käsitellään myös binäärilukuina, mutta ne esitetään luettavuuden parantamiseksi ja tilan säästämiseksi heksadesimaaleina.

1. Luodaan viesti M', jonka alkuun kopioidaan viesti M, sen perään ykkönen ja tarvittava määrä nollia sekä 64 bitin esitys viestin M pituudesta. Nollien määrä valitaan niin, että viestin M' pituus on $512k$, missä k on positiivinen kokonaisluku. Näin saatu viesti M' jaetaan 512 bitin lohkoihin $m_0, \ldots, m_{k-1}$.

2. Alustetaan apumuuttujat seuraavin luvuin

$$\begin{aligned} A_0 &= 0x67452301, \\ B_0 &= 0xEFCDAB89, \\ C_0 &= 0x98BADCFE, \\ D_0 &= 0x10325476 && \text{ja} \\ E_0 &= 0xC3D2E1F0. \end{aligned}$$

3. Määritellään apufunktio $f_j : \mathbb{B}_{32} \times \mathbb{B}_{32} \times \mathbb{B}_{32} \to \mathbb{B}_{32}$ seuraavalla tavalla kaikilla kokonaisluvuilla $j \in [0, 79]$

$$f_j(X, Y, Z) = \begin{cases} (X \wedge Y) \vee (X \wedge Z), & \text{kun } j \in [0, 19], \\ X \oplus Y \oplus Z, & \text{kun } j \in [20, 39] \text{ tai } j \in [60, 79] \text{ ja} \\ (X \wedge Y) \vee (X \wedge Z) \vee (Y \wedge Z), & \text{kun } j \in [40, 59], \end{cases}$$

missä $\wedge$ on bittikohtainen ja, $\vee$ on bittikohtainen inklusiivinen tai ja $\oplus$ on bittikohtainen eksklusiivinen tai.

Määritellään vakiot K_j seuraavalla tavalla kaikilla kokonaisluvuilla $j \in [0, 79]$

$$K_j = \begin{cases} 0x5A827999, & \text{kun } 0 \leq j \leq 19, \\ 0x6ED9EBA1 & \text{kun } 20 \leq j \leq 39, \\ 0x8F1BBCDC & \text{kun } 40 \leq j \leq 59, \\ 0xCA62C1D6 & \text{kun } 60 \leq j \leq 79. \end{cases}$$

4. Jokaisella kokonaisluvulla $i \in [0, k-1]$ suoritetaan seuraavat toiminnot.

   (a) Jaetaan kyseinen lohko m_i 32 bitin sanoiksi $w_0, \ldots, w_{15}$. Luodaan lisäksi kaikilla $j \in [16, 79]$ sanat

   $$w_j = w_{j-3} \oplus w_{j-8} \oplus w_{j-14} \oplus w_{j-16}.$$

   (b) Jokaisella kokonaisluvulla $j \in [0, 79]$ päivitetään apumuuttujat

   $$\begin{aligned} A_{j+1} &= w_j + rot_5(A_j) + f_j(B_j, C_j, D_j) + E_j + k_j \mod 2^{32}, \\ B_{j+1} &= A_j, \\ C_{j+1} &= rot_{30}(B_j), \\ D_{j+1} &= C_j && \text{ja} \\ E_{j+1} &= D_j, \end{aligned}$$

   missä $rot_n(X)$ kierrättää luvun X bittejä n paikkaa vasemmalle.

(c) Kun kyseinen lohko on käsitelty, niin alustetaan apumuuttujat seuraavaa lohkoa varten

$$\begin{aligned} A_0 &= A_0 + A_{80} \mod 2^{32}, \\ B_0 &= B_0 + B_{80} \mod 2^{32}, \\ C_0 &= C_0 + C_{80} \mod 2^{32}, \\ D_0 &= D_0 + D_{80} \mod 2^{32} \qquad \text{ja} \\ E_0 &= E_0 + E_{80} \mod 2^{32}. \end{aligned}$$

5. Kun kaikki lohkot on käsitelty, niin luodaan hasharvo $SHA(M)$, johon kopioidaan apumuuttujat järjestyksessä A_0, B_0, C_0, D_0, E_0 peräkkäin. Näin luotu hasharvo on $5 \times 32 = 160$ bittinen.

Täten määritelty SHA-0 on yksisuuntainen hashfunktio, mille ei ole kehitetty tehokasta tapaa löytää alkukuvia tai toisia alkukuvia. Sen sijaan Chabaud ja Joux osoittivat artikkelissaan [1], että SHA-0 ei ole törmäysresistantti.

Seuraavaksi suoritetaan lyhyt katsaus SHA-perheen historiaan Stinsonin [24, s. 139] ja Knospen [9, s. 142-149] kirjojen avulla. SHA tulee sanoista Secure Hash Algortihm, mikä kuvaa hyvin sitä, että kaikki SHA-funktiot ovat hashfunktioita. Vuonna 1993 julkaistu SHA-0, silloin pelkkä SHA, oli yhdysvaltalaisen National Institute of Standards and Technologyn (NIST) ensimmäinen standardisoitu hashfunktio. Hyvin pian kuitenkin alettiin epäillä, ettei se ole törmäysresistantti, ja vuonna 1995 se korvattiin SHA-1 funktiolla, jossa esimerkin 2.2 kohdan 4a sanojen luonti muutettiin seuraavanlaiseksi kaikilla kokonaisluvuilla $j \in [16, 79]$

$$w_j = rot_1(w_{j-3} \oplus w_{j-8} \oplus w_{j-14} \oplus w_{j-16}).$$

Aiemmin mainittu ranskalaisten Chabaud'n ja Joux'n artikkeli [1] julkaistiin vuonna 1998 osoittaen epäilyksen todeksi.

Vuonna 2002 NIST julkaisi uudet SHA-2 funktiot SHA-224, SHA-256, SHA-384 ja SHA-512, missä eri funktiot toimivat keskenään samalla periaatteella, mutta tuottavat eri mittaiset hasharvot. Sekä SHA-1 että SHA-2 funktiot on määritelty standardissa FIPS-180-4 [29] ja noudattavat useiden muiden hashfunktioiden tavoin Merkle-Damgård konstruktiota (ks. [24, s. 131-136]). Nykyään SHA-2 funktiot ovat laajalti käytössä ja muiden SHA-perheen funktioiden tapaan maailmanlaajuisen tutkimuksen kohteena tietoturvaominaisuuksiensa osalta.

NIST alkoi epäillä myös SHA-1 funktion törmäysresistanssia ja aloitti vuonna 2007 kilpailun uuden SHA-3 funktion kehittämiseksi. Useiden kierrosten jälkeen vuonna 2015 NIST julkaisi standardin FIPS 202 [30], joka määrittelee Keccak-algoritmiin perustuvat funktiot SHA3-224, SHA3-256, SHA3-384 ja SHA3-512 sekä SHAKE128 ja SHAKE256. Näistä SHA3 funktiot ovat hashfunktioita ja SHAKE-funktiot mahdollistavat myös eri mittaisten tulosten saamisen (extendable-output function, XOF). Molempia voidaan käyttää muiden SHA-perheen funktioiden tapaan hashfunktioina, vaikka SHAKE-funktioiden osalta tiettyä varovaisuutta on noudatettava.

Vuonna 2011 NIST lakkasi virallisesti suosittelemasta SHA-1 funktion käyttöä hashfunktiona. Vuonna 2017 alankomaalaiset Stevens ja Karpman yhdessä yhdysvaltalaisten Burszteinin, Albertinin ja Markov'n kanssa julkaisivat artikkelin [23], jossa he osoittavat SHA-1 funktion törmäysresistanssin murretuksi. SHA-2 ja SHA-3 funktioiden uskotaan yleisesti yhä olevan törmäysresistantteja.

2.3 Tietoturvaominaisuuksien riippuvuuksia

Edellisen alaluvun esimerkki 2.2 osoitti, että yksisuuntaisuus ei takaa, että hashfunktio on törmäysresistentti. Tässä alaluvussa keskitytäänkin tarkastelemaan, miten alkukuvaresistenssi ja toisen alkukuvan suhteen resistenssi riippuvat törmäysresistenssistä. Täten saadaan tietää, milloin hashfunktio on sekä yksisuuntainen että törmäysresistentti. Aloitetaan todistamalla, että törmäysresistenssi takaa, että hashfunktio on toisen alkukuvan suhteen resistentti.

Lause 2.1. *Vrt. [16, s. 329] ja [24, s. 127]. Jos hashfunktio on törmäysresistentti, niin se on toisen alkukuvan suhteen resistentti.*

Todistus. Olkoon $h\colon M \to \mathbb{B}_n$ törmäysresistentti hashfunktio. Tehdään vastaoletus, että hashfunktio h ei ole toisen alkukuvan suhteen resistentti. Valitaan jokin $m \in M$. Tällöin voidaan löytää määritelmän 2.6 kohdan 2 mukaan $m' \in M$, jolla $m \neq m'$ ja $h(m) = h(m')$. Täten on löydetty törmäys eli $m, m' \in M$, joilla $m \neq m'$ ja $h(m) = h(m')$, mikä on ristiriidassa törmäysresistanssin määritelmän 2.6 kohdan 4 kanssa. Näin ollen on osoitettu, että jos hashfunktio ei ole toisen alkukuvan suhteen resistentti, niin se ei ole törmäysresistentti. Tästä edelleen seuraa, että jos hashfunktio on törmäysresistentti, niin se on toisen alkukuvan suhteen resistentti. □

Seuraavassa lauseessa todistetaan, että minkä tahansa hashfunktion avulla voidaan jakaa mahdollisten viestien joukko ekvivalenssiluokkiin. Tätä tietoa hyödyntäen todistetaan sitten, että tietyissä tapauksissa törmäysresistenssi takaa myös alkukuvaresistenssin. Seurauksena saadaan muutamia mielenkiintoisia havaintoja liittyen siihen, miten törmäysresistenssi riippuu myös hashfunktion lähtö-, maali- ja arvojoukon koosta sekä viestien jakautumisesta eri ekvivalenssiluokkiin.

Lause 2.2. *Vrt. [24, s. 128]. Olkoon $h\colon M \to \mathbb{B}_n$ hashfunktio. Määritellään kaikilla viesteillä $m \in M$, että $m \sim m'$, jos $h(m) = h(m')$. Tällöin $\sim$ on ekvivalenssirelaatio. Merkitään viestin m määrittelemää ekvivalenssiluokkaa $[m]$ ja kaikkien ekvivalenssiluokkien joukkoa C.*

Todistus. Olkoot $h\colon M \to \mathbb{B}_n$ hashfunktio ja $m_0, m_1, m_3 \in M$. Määritellään kaikilla viesteillä $m \in M$, että $m \sim m'$, jos $h(m) = h(m')$. Tällöin ensinnäkin

$$m_0 \sim m_0, \text{ sillä } h(m_0) = h(m_0),$$

joten $\sim$ on refleksiivinen. Toiseksi, kun $m_0 \sim m_1$, niin

$$m_1 \sim m_0, \text{ sillä } h(m_1) = h(m_0) \text{ eli } h(m_0) = h(m_1),$$

joten $\sim$ on symmetrinen. Kolmanneksi, kun $m_0 \sim m_1$ ja $m_1 \sim m_2$, niin

$$m_0 \sim m_2, \text{ sillä } h(m_0) = h(m_1) = h(m_2),$$

joten $\sim$ on transitiviinen. Täten $\sim$ on ekvivalenssirelaatio, sillä se on refleksiivinen, symmetrinen ja transitiivinen. □

Lause 2.3. *Vrt. [24, s. 128-129]. Olkoon $h\colon M \to \mathbb{B}_n$ törmäysresistentti hashfunktio, jolle joukko M on äärellinen ja $|M| \geq 2|\mathbb{B}_n| = 2^{n+1}$. Tällöin hashfunktio h on alkukuvaresistentti.*

Todistus. Olkoon $h\colon M \to \mathbb{B}_n$ törmäysresistentti hashfunktio, jolle joukko M on äärellinen ja $|M| \geq 2^{n+1}$. Tehdään vastaoletus, että hashfunktio h ei ole alkukuvaresistantti. Olkoon $m \in M$, jolla $h(m) = b$. Tällöin voidaan löytää määritelmän 2.6 kohdan 1 mukaan $m' \in M$, jolla $h(m') = b$. Nyt todennäköisyys, että $m' \neq m$, ja täten on löydetty törmäys, on

$$\frac{|[m]| - 1}{|[m]|}.$$

Keskimääräinen todennäköisyys löytää törmäys riippumatta viestistä m on silloin

$$\frac{1}{|M|} \sum_{m \in M} \frac{|[m]| - 1}{|[m]|} = \frac{1}{|M|} \sum_{[m] \in C} \sum_{m \in [m]} \frac{|[m]| - 1}{|[m]|} \tag{2.1}$$

$$= \frac{1}{|M|} \sum_{[m] \in C} \frac{|[m]|(|[m]| - 1)}{|[m]|} = \frac{1}{|M|} \sum_{[m] \in C} (|[m]| - 1) \tag{2.2}$$

$$= \frac{1}{|M|} \sum_{[m] \in C} |[m]| - \sum_{[m] \in C} 1 \tag{2.3}$$

$$= \frac{|M| - |C|}{|M|}. \tag{2.4}$$

Yllä olevat yhtä suuruudet perustuvat seuraaviin havaintoihin.

- Rivillä 2.1 on hyödynnetty yhteenlaskun vaihdannaisuutta ja laskettu summa kaikkien viestien osalta ensin oman ekvivalenssiluokkansa sisällä ja sitten eri ekvivalenssiluokkien välillä.
- Rivillä 2.2 on muutettu sisempi summa kertolaskuksi, sillä kyseinen summa ei riipu viestistä itsestään, vaan sen ekvivalenssiluokan koosta, mikä on luonnollisesti samaan ekvivalenssiluokkaan kuuluvilla viesteillä yhtä suuri.
- Rivillä 2.3 on jälleen hyödynnetty yhteenlaskun vaihdannaisuutta ja jaettu rivillä 2.2 esiintyvä erotus $|[m]| - 1$, eli summa $|[m]| + (-1)$, kahteen erilliseen summaan.
- Rivillä 2.4 on hyödynnetty tietoa, että kaikkien ekvivalenssiluokkien kokojen summa on yhtä kuin kaikkien viestien joukon koko, sillä jokainen viesti kuuluu täsmälleen yhteen ekvivalenssiluokkaan. Lisäksi jälkimmäinen summa on muutettu kertolaskuksi $1|C|$, sillä sitä kyseinen summa tarkoittaa.

Kun tiedetään, että $|M| \geq 2^{n+1}$ ja ekvivalenssiluokkia on korkeintaan yksi jokaista hasharvoa kohti, eli $|C| \leq |\mathbb{B}_n| = 2^n$, niin voidaan arvioida todennäköisyyttä seuraavasti

$$\frac{|M| - |C|}{|M|} \geq \frac{2^{n+1} - 2^n}{2^{n+1}} = \frac{1}{2}.$$

Näin ollen keskimäärin vähintään puolessa tapauksista jo ensimmäisellä yrityksellä löytyy törmäys. Kun yrityksiä on enemmän, niin todennäköisyys löytää jollain niistä törmäys kasvaa edelleen. On mahdollista, mutta epätodennäköistä, että törmäystä ei löydy. Kuitenkin on todistettu, että ei välttämättä ole vaikeaa löytää $m, m' \in M$, joilla $m \neq m'$ ja $h(m) = h(m')$, mikä on ristiriidassa törmäysresistenssin määritelmän 2.6 kohdan 4 kanssa. □

Seuraus 2.4. *Olkoot $h\colon M \to \mathbb{B}_n$ hashfunktio ja Y sen arvojoukko, jossa kaikilla $m \in M$ pätee $h(m) \in Y$. Tällöin todennäköisyys löytää törmäys kasvaa seuraavissa tapauksissa.*

1. *Lähtöjoukon M koko kasvaa suhteessa maalijoukkoon $\mathbb{B}_n$.*
2. *Maalijoukon $\mathbb{B}_n$ koko pienenee suhteessa lähtöjoukkoon M eli hasharvot lyhenevät.*
3. *Arvojoukon Y koko pienenee suhteessa maalijoukkoon $\mathbb{B}_n$.*
4. *Viestit eivät jakaudu tasaisesti eri ekvivalenssiluokkiin.*

Todistus. Olkoot $h\colon M \to \mathbb{B}_n$ hashfunktio ja Y sen arvojoukko, jossa kaikilla $m \in M$ pätee $h(m) \in Y$. Tarkastellaan lauseen 2.3 todistuksen kaavarivin 2.4 todennäköisyyttä $\frac{|M|-|C|}{|M|}$.

1. Kun joukon M koko kasvaa suhteessa joukkoon $\mathbb{B}_n$, niin joukon C koko pysyy vakiona ja
$$\lim_{|M|\to\infty} \frac{|M| - |C|}{|M|} = 1.$$
2. Kun joukon $\mathbb{B}_n$ koko pienenee suhteessa joukkoon M, niin joukon C koko pienenee ja
$$\lim_{|C|\to 0} \frac{|M| - |C|}{|M|} = 1.$$
3. Kun joukon Y koko pienenee suhteessa joukkoon $\mathbb{B}_n$, niin joukon C koko pienenee, sillä ekvivalenssiluokkia on täsmälleen yksi jokaista hasharvoa eli joukon Y alkiota kohti. Tällöin todennäköisyys löytää törmäys lähestyy yhtä, kuten edellisessä kohdassa.
4. Tarkastellaan nyt todennäköisyyttä $\frac{|[m]|-1}{|[m]|}$. Kun viestit eivät jakaudu tasaisesti eri ekvivalenssiluokkiin, niin jokin ekvivalenssiluokka on muita isompi ja siten on suurin todennäköisyys, että siihen kuuluu lauseen 2.3 todistuksen alussa kiinnitetty $m \in M$, jolla $h(m) = b$. Tällöin
$$\lim_{|[m]|\to|M|} \frac{|[m]| - 1}{|[m]|} \approx 1.$$

Näin ollen kaikissa edellä mainituissa tapauksissa todennäköisyys löytää törmäys kasvaa. □

Tämän alaluvun lopuksi määritellään vielä todistettavissa oleva hashfunktio, jonka törmäysresistenssi riippuu jostakin matemaattisesta ongelmasta.

Määritelmä 2.7. Vrt. [17, s. 43] ja [21, s. 210]. Olkoon h hashfunktio. Kun hashfunktion h törmäysresistenssi perustuu johonkin vaikeaksi tiedettyyn tai oletettuun matemaattiseen ongelmaan, niin h on todistettavissa oleva (provable) hashfunktio.

Vaikeaksi tiedetyillä matemaattisilla ongelmilla tarkoitetaan tässä NP-kovia ongelmia ja vaikeaksi oletetuilla paljon tutkittuja, mutta ratkaisemattomia ongelmia, kuten alkulukutekijöihin jako ja diskreettilogaritmi.

Esimerkki 2.3. Vrt. [22, luku 18.12]. Olkoot p ja q alkulukuja sekä $k > 1$ kokonaisluku, joka on keskenään jaoton luvun $(p-1)(q-1)$ kanssa. Määritellään hashfunktio $h\colon M \to \mathbb{B}_n$, missä $n \geq \log_2 pq$, seuraavasti kaikilla mahdollisilla viesteillä $m \in M$

$$h(m) = m^k \mod pq.$$

Hashfunktion h määritelmän johdosta sen törmäysresistenssi perustuu RSA-salausmenetelmän tavoin moduluksen pq alkulukutekijöihin jakoon, mille ei ole kehitetty tehokasta laskentatapaa. Täten hashfunktio h on määritelmän 2.7 mukaan todistettavissa oleva.

Sen sijaan esimerkin 2.2 hashfunktio SHA-0 ei ole todistettavissa oleva.

2.4 Väsytyshyökkäyksiä hashfunktioita vastaan

Väsytyshyökkäykset (brute-force attack) perustuvat siihen, että kokeillaan mahdollisia ratkaisuja, kunnes löydetään haluttu ratkaisu. Niiden ei katsota murtavan hashfunktion tietoturvaominaisuuksia, mutta ne asettavat tiettyjä vähimmäisvaatimuksia hasharvojen pituudelle. Aloitetaan määrittelemällä tyhjentävä haku (exhaustive search), joka on hashfunktion alkukuvaresistenssia tai toisen alkukuvan suhteen resistenssia vastaan suunnattu väsytyshyökkäys.

Määritelmä 2.8. Vrt. [17, s. 29] ja [22, luku 7.4.]. Olkoot $h\colon M \to \mathbb{B}_n$ hashfunktio ja $b \in \mathbb{B}_n$ jokin sen tuottama hasharvo. Tyhjentävässä haussa alkukuvaresistenssia vastaan hyökkääjä muodostaa listan $m_0, m_1, m_2, \cdots \in M$ ja laskee siinä järjestyksessä viesteille hasharvoja verraten niitä aina hasharvoon b, kunnes löytää viestin m_i, jolla $h(m_i) = b$.

Määritelmä 2.9. Vrt. [17, s. 29] ja [22, luku 7.4.]. Olkoot $h\colon M \to \mathbb{B}_n$ hashfunktio ja $m \in M$. Tyhjentävässä haussa toisen alkukuvan suhteen resistenssia vastaan hyökkääjä muodostaa listan $m_0, m_1, m_2, \cdots \in M \setminus \{m\}$ ja laskee siinä järjestyksessä viesteille hasharvoja verraten niitä aina hasharvoon $h(m)$, kunnes löytää viestin m_i, jolla $h(m_i) = h(m)$.

Lause 2.5. *Vrt. [17, s. 29] ja [22, luku 7.4.]. Olkoon $h\colon M \to \mathbb{B}_n$ hashfunktio. Tyhjentävällä haulla löydetään alkukuva tai toinen alkukuva 2^n kokeilulla.*

Todistus. Olkoot $h\colon M \to \mathbb{B}_n$ hashfunktio ja Y sen arvojoukko.

Tarkastellaan ensin todennäköisyyttä löytää alkukuva hasharvolle $b \in Y$. Valitaan $m \in M$. Tällöin todennäköisyys, että $h(m) = b$ ja täten on löydetty alkukuva, on

$$\frac{|[b^{-1}]|}{|M|},$$

missä b^{-1} on jokin hasharvon b alkukuva. Keskimääräinen todennäköisyys löytää alkukuva riippumatta hasharvosta b on silloin

$$\begin{aligned}\frac{1}{|Y|}\sum_{b\in Y}\frac{|[b^{-1}]|}{|M|} &= \frac{1}{|Y||M|}\sum_{b\in Y}|[b^{-1}]| \\ &= \frac{|M|}{|Y||M|} = \frac{1}{|Y|} \\ &\geq \frac{1}{2^n}.\end{aligned}$$

Näin ollen todennäköisyys löytää alkukuva 2^n kokeilulla on vähintään

$$2^n\frac{1}{2^n} = 1.$$

Tarkastellaan sitten todennäköisyyttä löytää toinen alkukuva viestille $m \in M$. Käytännössä se vastaa alkukuvan löytämistä joukosta $M \setminus \{m\}$. Merkitään $h(m) = b$ ja valitaan $m' \in M \setminus \{m\}$. Tällöin todennäköisyys, että $h(m') = b$ ja täten on löydetty toinen alkukuva, on

$$\frac{|[b^{-1}] \setminus \{m\}|}{|M \setminus \{m\}|}.$$

Keskimääräinen todennäköisyys löytää alkukuva riippumatta viestistä m on silloin

$$\begin{aligned}\frac{1}{|Y|}\sum_{b\in Y}\frac{|[b^{-1}] \setminus \{m\}|}{|M \setminus \{m\}|} &= \frac{1}{|Y|(|M|-1)}\sum_{b\in Y}|[b^{-1}] \setminus \{m\}| \\ &= \frac{|M|-1}{|Y|(|M|-1)} = \frac{1}{|Y|} \\ &\geq \frac{1}{2^n}.\end{aligned}$$

Näin ollen todennäköisyys löytää toinen alkukuva 2^n kokeilulla on myös vähintään 1. Ja täten tyhjentävällä haulla löydetään alkukuva tai toinen alkukuva 2^n kokeilulla. □

Seuraavaksi hyökätään hashfunktion törmäysresistanssia vastaan. Yleisesti tämä väsytyshyökkäys tunnetaan syntymäpäivähyökkäyksenä (birthday attack), sillä se perustuu syntymäpäiväparadoksiin. Syntymäpäiväparadoksihan on, että 23 hengen joukossa todennäköisyys on suurempi kuin $0,5$, että kahdella henkilöllä on sama syntymäpäivä.

Määritelmä 2.10. Vrt. [17, s. 29] ja [22, luku 7.4 ja luku 18.1]. Olkoon $h\colon M \to \mathbb{B}_n$ hashfunktio. Syntymäpäivähyökkäyksessä hyökkääjä muodostaa listan $m_0, m_1, m_2, \cdots \in M$ ja laskee siinä järjestyksessä viesteille hasharvoja verraten niitä aina kaikkiin aiemmin laskemiinsa hasharvoihin, kunnes löytää viestit m_i ja m_j, joilla $h(m_i) = h(m_j)$.

Lause 2.6. *Vrt. [17, s. 29] ja [22, luku 7.4.]. Olkoon $h\colon M \to \mathbb{B}_n$ hashfunktio. Syntymäpäivähyökkäyksellä löydetään törmäys $2^{(n+1)/2}$ kokeilulla.*

Todistus. Olkoot $h\colon M \to \mathbb{B}_n$ hashfunktio, Y sen arvojoukko ja $m_0, m_1, m_2, \cdots \in M$.

Todennäköisyys löytää törmäys viestillä m_1 vastaa käytännössä toisen alkukuvan löytämistä viestille m_0. Tämä on lauseen 2.5 todistuksen perusteella $\frac{1}{2^n}$. Vastaavasti todennäköisyys löytää törmäys viestillä m_2 vastaa käytännössä toisen alkukuvan löytämistä viestille m_0 tai m_1. Tämä on $\frac{2}{2^n}$. Edelleen todennäköisyys löytää törmäys viestillä m_j vastaa käytännössä toisen alkukuvan löytämistä jollekin viesteistä $m_0, \ldots, m_{j-1}$. Tämä on $\frac{j}{2^n}$.

Näin ollen todennäköisyys löytää törmäys jollakin viesteistä $m_1, \ldots, m_j$ on

$$\sum_{i=1}^{j} \frac{i}{2^n} = \frac{1}{2^n} \sum_{i=1}^{j} i = \frac{1}{2^n} \frac{i(i+1)}{2} = \frac{i(i+1)}{2^{n+1}}.$$

Kun kokeiluja on $2^{(n+1)/2}$, niin $i = 2^{(n+1)/2}$, ja saadaan todennäköisyys

$$\frac{2^{(n+1)/2}(2^{(n+1)/2}+1)}{2^{n+1}} = \frac{2^{n+1}+2^{(n+1)/2}}{2^{n+1}} = 1 + \frac{1}{2^{(n+1)/2}} > 1.$$

Näin ollen todennäköisyys löytää törmäys $2^{(n+1)/2}$ kokeilulla on suurempi kuin 1. Ja täten syntymäpäivähyökkäyksellä löydetään törmäys $2^{(n+1)/2}$ kokeilulla. □

Tässä alaluvussa esitettyjen väsytyshyökkäysten johdosta Schneier [22] suosittelee käyttämään hashfunktioita, jotka tuottavat vähintään 160 bittisiä hasharvoja. Kirjan julkaisun jälkeen hashfunktion SHA-1 törmäysresistanssi on murrettu huolimatta 160 bittisistä hasharvoista. NIST suositteleekin korvaamaan sen jollakin SHA-2 tai SHA-3 funktioista, jotka tuottavat lyhimmillään 224 bittisiä hasharvoja.

3 Graafiteorian perusteet

Tässä luvussa tarkastellaan yleisesti graafeja ja luodaan perusteet ymmärtää luvussa 4 esitettäviä laajenevien graafien perheitä (expander families/graphs). Graafit ovat helposti ymmärrettävissä paperille piirrettyinä palloina eli solmuina, joita viivat eli särmät yhdistävät toisiinsa. Niiden avulla pystytäänkin mallintamaan luontevasti erilaisia relaatioita ja esimerkiksi tietokoneista muodostettuja verkkoja. Graafiteoria tunnetaankin myös verkkoteoriana.

3.1 Graafit, solmut ja särmät

Aloitetaan aivan alusta määrittelemällä graafi, solmu ja särmä sekä esittämällä niille erilaisia merkintätapoja. Esimerkissä piirretään graafi paperille palloina ja viivoina.

Määritelmä 3.1. Vrt. [4, s. 2] ja [6, luku 1.1.1]. Olkoot V ja E äärellisiä joukkoja. Jos jokaisella $\{a, b\} \in E$ pätee $a, b \in V$, niin $G = (V, E)$ on (suuntaamaton) graafi. Joukon V alkiot ovat solmuja ja joukon E alkiot ovat (suuntaamattomia) särmiä.

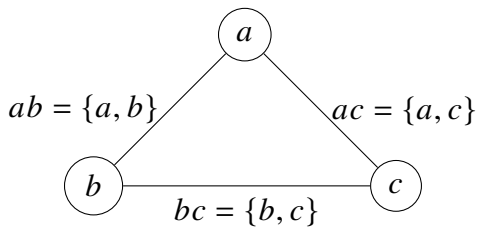
Määritelmä 3.2. Vrt. [4, s. 2] ja [6, luku 1.1.1]. Olkoon G graafi.

1. Graafin G kaikkien solmujen joukkoa voidaan merkitä $V(G)$ ja kaikkien särmien joukkoa $E(G)$, vaikka niitä ei olisi nimetty joukoiksi V ja E.

2. Särmä $\{a, b\}$ voidaan merkitä ilman joukkomerkintää ab.

3. Voidaan merkitä $a \in G$ ja $ab \in G$ tarkoittamaan, että $a \in V(G)$ ja $ab \in E(G)$.

Esimerkki 3.1. Vrt. [4, s. 2], [6, luku 1.1.1] ja [13, s. 8–11]. Kun $V = \{a, b, c\}$ ja $E = \{\{a, b\}, \{a, c\}, \{b, c\}\}$, niin graafi $G = (V, E)$ voidaan esittää seuraavalla tavalla.



Kuva 3.1: Graafi G esitettynä kuvana

Seuraavaksi määritellään muutamia solmujen ja särmien yleisimpiä ominaisuuksia. Niitä tarvitaan, kun määritellään yksinkertainen graafi sekä nolla- ja triviaaligraafi, jotka ovat pienimpiä mahdollisia graafeja.

Määritelmä 3.3. Vrt. [4, s. 2–3, 28] ja [6, luku 1.1.1]. Olkoon G graafi, jossa solmujen a ja b välillä on särmä ab. Tällöin käytetään seuraavia termejä.

1. **Päätepiste:** Solmut a ja b ovat särmän ab päätepisteet.
2. **Vierussolmu:** Solmut a ja b ovat toistensa vierussolmut.
3. **Luuppi:** Jos $a = b$, eli $ab = aa$, niin se on luuppi (loop), ja solmu on itsensä vierussolmu.
4. **Moninkertainen särmä:** Moninkertainen särmä on joukko, johon kuuluu vähintään kaksi särmää, joilla on samat päätepisteet.

Määritelmä 3.4. Vrt. [6, luku 1.1.1]. Yksinkertainen graafi on graafi, jossa ei ole yhtään luuppia tai moninkertaista särmää.

Esimerkki 3.2. Esimerkin 3.1 graafi G on yksinkertainen graafi, sillä siinä ei ole luuppeja ja siinä on vain yksittäisiä särmiä solmujen välillä.

Määritelmä 3.5. Vrt. [4, s. 2] ja [6, luku 1.1.1]. Nollagraafi on graafi, jossa ei ole yhtään solmua tai särmää.

Määritelmä 3.6. Vrt. [4, s. 2] ja [6, luku 1.1.1]. Triviaaligraafi on graafi, jossa on yksi solmu ja ei yhtään särmää.

Näiden suuntaamattomien graafien avulla voidaan mallintaa vain symmetrisiä relaatioita. Epäsymmetristen relaatioiden mallintamiseen tarvitaan suunnattuja graafeja, joita tarkastellaan seuraavaksi. Ensin määritellään suunnattu graafi ja sen jälkeen osoitetaan, että jokainen suuntaamaton graafi voidaan esittää suunnattuna graafina sekä esitellään graafin suuntiminen.

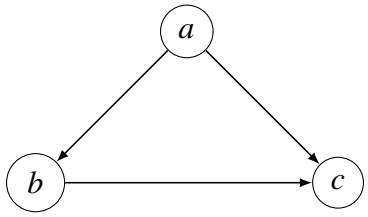
Määritelmä 3.7. Vrt. [4, s. 28] ja [6, luku 1.1.1]. Olkoot V ja E äärellisiä joukkoja. Jos jokaisella $(a, b) \in E$ pätee $a, b \in V$, niin $G = (V, E)$ on suunnattu graafi. Joukon E alkiot ovat suunnattuja särmiä, joita voidaan merkitä ilman parimerkintää $\overline{ab}$.

Lause 3.1. *Jokainen suuntaamaton graafi voidaan esittää suunnattuna graafina.*

Todistus. Olkoon $G = (V, E)$ suuntaamaton graafi. Luodaan suunnattujen särmien joukko E' siten, että jokaisella $ab \in E$ pätee $\overline{ab}, \overline{ba} \in E'$. Tällöin $G' = (V, E')$ on G esitettynä suunnattuna graafina. □

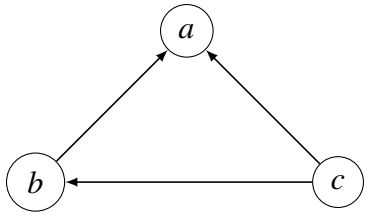
Määritelmä 3.8. Vrt. [10, s. 20]. Olkoon $G = (V, E)$ suuntaamaton graafi. Graafi G suunnitaan korvaamalla jokainen $ab \in E$ asettamalla joko $\overline{ab} \in E$ tai $\overline{ba} \in E$. Lisäksi jokaista $\overline{e} = \overline{ab} \in E$ kohti merkitään $a = e^-$ ja $b = e^+$.

Esimerkki 3.3. Vrt. [13, s. 4]. Esimerkin 3.1 graafi G voidaan suuntia asettamalla $E = \{\overline{ab}, \overline{ac}, \overline{bc}\}$. Alla on esitetty graafi G suunnittuna.



Kuva 3.2: Suunnittu graafi G

Huomataan myös, että graafin suuntiminen ei ole yksikäsitteinen, vaan myös graafi G' on graafi G suunnittuna.



Kuva 3.3: Graafi G'

Jatkossa kuitenkin graafit ovat aina suuntaamattomia, ellei toisin mainita. Tämän alaluvun lopuksi määritellään aligraafi, joka luo tietyllä tavalla samankaltaisten graafien välille yhteyden, sekä täydelllinen graafi.

Määritelmä 3.9. Vrt. [4, s. 4] ja [28, s. 6]. Olkoot G ja H graafeja. Jos $V(H) \subseteq V(G)$ ja $E(H) \subseteq E(G)$, niin H on graafin G aligraafi.

Esimerkki 3.4. Esimerkin 3.1 graafin G joitakin aligraafeja ovat triviaaligraafi $(\{a\}, \emptyset)$, graafi $H = (\{a, b\}, \{ab\})$ sekä graafi G itse.

Määritelmä 3.10. Vrt. [4, s. 3] ja [28, s. 9]. Olkoon G yksinkertainen graafi, jonka kaikki solmut ovat toistensa vierussolmuja. Tällöin G on täydellinen graafi.

3.2 Solmun ja graafin aste sekä säännöllisyys

Siirrytään tarkastelemaan asteita. Aloitetaan solmun asteesta, jolla saadaan mielekkäällä tavalla kuvattua särmien määrää kutakin solmua kohti. Käytännössä siis lasketaan paperille piirretystä pallosta siihen yhdistyvät viivat. Määritellään myös yksittäinen solmu, joka ei ole minkään särmän päätepiste.

Määritelmä 3.11. Vrt. [4, s. 5 ja 28–29], [6, luku 1.1.2] ja [28, s. 34]. Solmun a aste, $\deg(a)$, on summa luuppien aa ja niiden särmien, joilla on päätepiste solmussa a, lukumäärästä.

Määritelmä 3.12. Vrt. [4, s. 5] ja [6, luku 1.1.2]. Jos $\deg(a) = 0$, niin a on yksittäinen solmu.

Seuraavaksi laajennetaan ajattelu asteista koskemaan kokonaista graafia ja siellä solmujen ja särmien lukumäärän suhteeseen. Määritellään graafille minimi-, maksimi- ja keskimääräinen aste sekä havainnollistetaan asteita esimerkillä.

Määritelmä 3.13. Vrt. [4, s. 5] ja [28, s. 34]. Graafin G minimiaste, $\delta(G)$, on pienin aste sen solmujen joukossa eli

$$\delta(G) = \min_{v \in V(G)} \deg(v).$$

Vastaavasti graafin G maksimiaste, $\Delta(G)$, on suurin aste sen solmujen joukossa eli

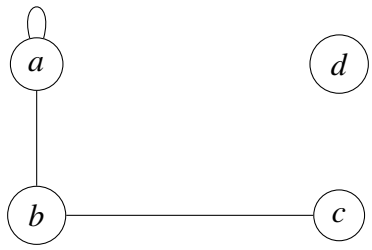
$$\Delta(G) = \max_{v \in V(G)} \deg(v).$$

Määritelmä 3.14. Vrt. [4, s. 5] ja [28, s. 35]. Graafin G keskimääräinen aste, $\deg(G)$, on sen solmujen asteiden keskiarvo eli

$$\deg(G) = \frac{1}{|V(G)|} \sum_{v \in G} \deg(v),$$

missä $|V(G)|$ on joukon $V(G)$ koko eli solmujen määrä graafissa G.

Esimerkki 3.5. Vrt. [13, s. 8–11]. Olkoon H alla esitetty graafi.



Kuva 3.4: Graafi H

Tällöin $\deg(a) = 3, \deg(b) = 2, \deg(c) = 1$ ja $\deg(d) = 0$ eli d on yksittäinen solmu. Lisäksi minimiaste on $\delta(H) = 0$, maksimiaste on $\Delta(H) = 3$ ja keskimääräinen aste on $\deg(H) = \frac{3+2+1+0}{4} = 1,5$.

Seuraava lause tunnetaan yleisesti nimellä kättelylemma. Siinä osoitetaan, että asteiden ja särmien lukumäärän välillä todella on vakioitu yhteys. Tämän avulla saadaan myös osoitettua, että paritonta astetta olevia solmuja ei voi olla paritonta määrää missään graafissa.

Lause 3.2. *Vrt. [4, s. 5] ja [28, s. 35]. Olkoon G graafi. Tällöin*

$$\sum_{v \in G} \deg(v) = 2|E(G)|.$$

Todistus. Jokainen graafin G särmä lasketaan sen molemmissa päätepisteissä kyseisen solmun asteeseen ja siten kunkin särmän olemassa olo korottaa summaa

$$\sum_{v \in G} \deg(v)$$

kahdella. Erityisesti luupit lasketaan ensin kerran päätepisteenä ja toisen kerran määritelmän 3.11 mukaan.Täten

$$\sum_{v \in G} \deg(v) = 2|E(G)|.$$

□

Lause 3.3. *Vrt. [4, s. 5] ja [28, s. 35]. Paritonta astetta olevia solmuja on parillinen määrä kaikissa graafeissa.*

Todistus. Tehdään vastaoletus, että on olemassa graafi G, jossa on pariton määrä paritonta astetta olevia solmuja. Tällöin

$$\sum_{d \in G} \deg(v)$$

on pariton, mikä on ristiriidassa lauseen 3.2 kanssa, sillä $|E(G)|$ on aina kokonaisluku. □

Vielä määritellään asteiden avulla säännölliset graafit, jotka ovat erityisen tarpeellisia laajenevien graafien perheissä.

Määritelmä 3.15. Vrt. [4, s. 5] ja [28, s. 34]. Olkoon G graafi. Jos kaikkien graafin G solmujen aste on sama, niin G on säännöllinen. Kun yhteinen aste on d, niin voidaan täsmentää, että G on d-säännöllinen.

Esimerkki 3.6. Esimerkin 3.5 graafi H ei ole säännöllinen. Sen sijaan esimerkin 3.1 graafi G on 2-säännöllinen.

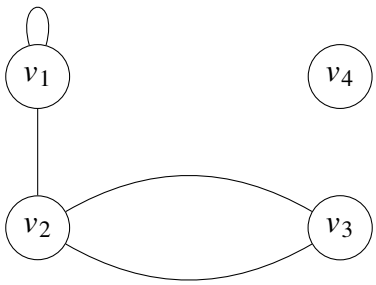
3.3 Spektraaligraafiteorian perusteita

Tässä alaluvussa perehdytään spektraaligraafiteorian perusteisiin, missä yhdistellään perinteiseen graafiteoriaan lineaarialgebraa. Käsiteltävät asiat ovat merkittävässä roolissa laajenevien graafien perheissä. Näistä ensimmäisenä määritellään vierusmatriisi ja sen avulla ominaisarvot. Asioita havainnollistetaan esimerkillä.

Määritelmä 3.16. Vrt. [10, s. 11] ja [28, s. 6]. Olkoon $G = (V, E)$ graafi, missä $V = \{v_1, v_2, \ldots, v_n\}$. Graafin G vierusmatriisi on $n \times n$-matriisi A, missä kaikilla $i, j \leq n$ alkio a_{ij} kertoo särmien $v_i v_j$ lukumäärän joukossa E.

Määritelmä 3.17. Vrt. [10, s. 12] ja [28, s. 453]. Olkoot G graafi ja A sen vierusmatriisi. Vierusmatriisin ominaisarvot ovat sen karakteristisen polynomin, $\det(A - xI)$, juuret $\lambda_0, \lambda_1, \ldots, \lambda_{n-1}$. Graafin G ominaisarvot ovat sen vierusmatriisin ominaisarvot. Merkitään graafin G ominaisrvoja $\lambda_0(G), \lambda_1(G), \ldots, \lambda_{n-1}(G)$.

Esimerkki 3.7. Laskutoimituksissa on hyödynnetty sovellusta [31]. Olkoon $G = (V, E)$ alla esitetty graafi.



Kuva 3.5: Graafi G

Graafin G vierusmatriisi on

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Matriisin A karakteristinen polynomi on $\lambda^4 - \lambda^3 - 4\lambda^2 + 2\lambda$, jonka juuret ja samalla graafin G ominaisarvot ovat $\lambda_0 \approx 2,34$, $\lambda_1 \approx 0,47$, $\lambda_2 = 0$ ja $\lambda_3 \approx -1,81$.

Jatkossa on oleellista, että ominaisarvot voidaan järjestää suurimmasta pienimpään. Tämä on mahdollista koska suuntaamattoman graafin vierusmatriisi on symmetrinen ja sen kaikki alkiot ovat reaalilukuja, mistä Westin [28] mukaan seuraa, että myös sen ominaisarvot ovat reaalilukuja. Esitetään nämä tulokset lauseina.

Lause 3.4. *Vrt. [10, s.12 ja 239] ja [28, s. 456]. Olkoon G suuntaamaton graafi. Tällöin $\lambda_0(G), \lambda_1(G), \ldots, \lambda_{n-1}(G) \in \mathbb{R}$.*

Seuraus 3.5. *Vrt. [10, s. 12]. Graafin G ominaisarvot voidaan järjestää seuraavasti*

$$\lambda_0(G) \geq \lambda_1(G) \geq \cdots \geq \lambda_{n-1}(G).$$

Seuraavaksi muodostetaan näin saaduista ominaisarvoista graafille spektri ja määritellään säännöllisille graafeille spektrin aukon koko (spectral gap).

Määritelmä 3.18. Vrt. [8, s. 453] ja [10, s. 12]. Olkoot G graafi ja $\Lambda = \{\lambda_0(G), \lambda_1(G), \ldots, \lambda_{n-1}(G)\}$ sen ominaisarvot. Graafin G spektri on

$$S = \begin{bmatrix} \lambda_0 & \lambda_1 & \cdots & \lambda_r \\ m_0 & m_2 & \cdots & m_r \end{bmatrix},$$

missä kaikilla $r \leq n$ pätee $\lambda_0, \lambda_1, \cdots, \lambda_r \in \Lambda$ ja $\lambda_0 \neq \lambda_1 \neq \cdots \neq \lambda_r$. Lisäksi kaikilla $i \leq r$ luvut m_i kertovat kuinka monesti kyseinen λ_i esiintyy joukossa Λ.

Määritelmä 3.19. Vrt. [8, s 454] ja [10, s. 32]. Olkoot G d-säännöllinen graafi ja $\lambda_1(G)$ sen toiseksi suurin ominaisarvo. Spektrin aukon koko (spectral gap) on $d - \lambda_1(G)$.

Jatkossa merkitään $F(S, \mathbb{C})$ joukolle, joka sisältää kaikki funktiot $f\colon S \to \mathbb{C}$, missä S on jokin äärellinen joukko. Tämä merkintä helpottaa seuraavaa määritelmää ja esimerkkiä, joissa vierusmatriisista jalostetaan vierusoperaattori.

Määritelmä 3.20. Vrt. [10, s. 14]. Olkoot $G = (V, E)$ graafi ja A sen vierusmatriisi sekä $f \in F(V, \mathbb{C})$. Graafin G vierusoperaattori, $A\colon F(V, \mathbb{C}) \to F(V, \mathbb{C})$, on kaikilla $v_k \in G$

$$(Af)(v_k) = \sum_{j=1}^{n} a_{k,j} f(v_j),$$

missä kaikki $a_{k,j}$ ovat vierusmatriisin A alkioita.

Esimerkki 3.8. Olkoot graafi G kuten esimerkissä 3.7 ja $f \in F(V, \mathbb{C})$, jolla kaikilla $v_k \in V$ pätee $f(v_k) = k$. Tällöin vierusoperaattori, A, on

$$(Af)(v_k) = \sum_{j=1}^{4} a_{k,j} f(v_j) = \sum_{j=1}^{4} a_{k,j} \cdot j.$$

Näin saadaan esimerkiksi

$$(Af)(v_1) = \sum_{j=1}^{4} a_{1,j} j = 1 \cdot 1 + 1 \cdot 2 + 0 \cdot 3 + 0 \cdot 4 = 3.$$

Nähdään, että käytännössä vierusoperaattori siis summaa yhteen kyseisestä solmusta v_k lähtevät särmät painottaen niitä funktion f arvolla toisessa päätepisteessä ja luuppien osalta painoarvona on $f(v_k)$.

Vierusmatriisi ja -operaattori ovat itsessäänkin tärkeitä, mutta niiden avulla saadaan myös määriteltyä Laplacen matriisi ja operaattori. Ensiksi tarkastellaan Laplacen matriisia ja sen käyttäytymistä säännöllisillä graafeilla.

Määritelmä 3.21. Vrt. [17, s. 75] ja [28, s. 462-463]. Olkoot $G = (V, E)$ graafi, missä $V = \{v_1, v_2, \cdots, v_n\}$, ja A sen vierusmatriisi. Olkoon D $n \times n$-matriisi, missä kaikilla $i \leq n$ alkio $d_{ii} = \deg(v_i)$ ja kaikilla $i \neq j$ alkio $d_{ij} = 0$. Graafin G Laplacen matriisi on $L = D - A$.

Esimerkki 3.9. Olkoon graafi G kuten esimerkissä 3.7. Graafin G Laplacen matriisi on

$$L = D - A = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 3 & -2 & 0 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Lause 3.6. *Vrt. [17, s. 75]. Olkoot G d-säännöllinen graafi, jolle $|V(G)| = n$, ja A sen vierusmatriisi. Tällöin graafin G Laplacen matriisi on $L = dI - A$, missä I on $n \times n$ yksikkömatriisi.*

Todistus. Olkoot G d-säännöllinen graafi ja A sen vierusmatriisi. Tällöin kaikilla $v_n \in V(G)$ $\deg(v_n) = d$ ja $D = dI$. Täten graafin G Laplacen matriisi on $L = dI - A$. □

Seuraavassa määritelmässä Laplacen operaattoria varten graafi täytyy suuntia, mutta suunnittaessa valituista särmien suunnista huolimatta lopputulos on kuitenkin aina sama. Tämä voidaan osoittaa säännöllisten graafien osalta kiertämällä suuntiminen ja hyödyntämällä sen sijaan kaikkien solmujen yhteistä astetta ja vierusoperaattoria.

Määritelmä 3.22. Vrt, [10, s. 21]. Olkoon $G = (V, E)$ graafi. Suunnitaan G ja määritellään $\delta\colon F(V, \mathbb{C}) \to F(E, \mathbb{C})$ ja $\delta^*\colon F(E, \mathbb{C}) \to F(V, \mathbb{C})$ kaikilla $f \in F(V, \mathbb{C})$ ja $g \in F(E, \mathbb{C})$ seuraavasti

$$(\delta f)(\overline{e}) = f(e^+) - f(e^-) \text{ ja } (\delta^* g)(v) = \sum_{\overline{e} \in E, v = e^+} g(\overline{e}) - \sum_{\overline{e} \in E, v = e^-} g(\overline{e}).$$

Graafin G Laplacen operaattori, $\Delta\colon F(V, \mathbb{C}) \to F(V, \mathbb{C})$, on $\Delta = \delta^* \delta$.

Lause 3.7. *Vrt. [10, s. 22]. Olkoot $G = (V, E)$ d-säännöllinen graafi ja A sen vierusoperaattori. Tällöin graafin G Laplacen operaattori on $\Delta = dI - A$.*

Todistus. Olkoot $G = (V, E)$ d-säännöllinen graafi ja A sen vierusoperaattori. Olkoot $f \in F(V, \mathbb{C})$, $v_k \in V$ ja $|V| = n$. Tällöin määritelmän 3.22 mukaan

$$\begin{aligned}
(\Delta f)(v_k) &= (\delta^*(\delta f))(v_k) \\
&= \sum_{\overline{e} \in E, v_k = e^+} (\delta f)(\overline{e}) - \sum_{\overline{e} \in E, v_k = e^-} (\delta f)(\overline{e}) \\
&= \left(\sum_{\overline{e} \in E, v_k = e^+} f(v_k) - \sum_{\overline{e} \in E, v_k = e^+, v_j = e^-} f(v_j) \right) \\
&- \left(\sum_{\overline{e} \in E, v_k = e^-, v_j = e^+} f(v_j) - \sum_{\overline{e} \in E, v_k = e^-} f(v_k) \right) \\
&= df(v_k) - \sum_{j=1}^{n} a_{k,j} f(v_j) \\
&= ((dI - A)f)(v_k).
\end{aligned}$$

□

3.4 Erilaisia polkuja graafissa

Nyt käsitellään kulkemista graafissa. Ajatellaan solmuja paikkoina, joissa halutaan käydä, ja särmiä teinä niiden välillä. Erilaiset reitit, joissa paikat ja tiet vuorottelevat, ovat polkuja (walk). Niihin liittyy aina jokin lähtöpaikka eli alkusolmu, määränpää eli loppusolmu ja matkan pituus. Määritellään erityyppisiä polkuja, annetaan niille pituus ja huomataan, että osa niistä voidaan esittää myös yksinkertaisena graafina.

Määritelmä 3.23. Vrt. [4, s. 10], [6, luku 1.1.2] ja [28, s. 20]. Olkoot G graafi ja $v_0, \ldots, v_n \in V(G)$ siten, että $v_0 v_1, \ldots, v_{n-1} v_n \in E(G)$. Merkkijono
$W = v_0, v_0 v_1, v_1, v_1 v_2, \ldots, v_{n-1} v_n, v_n$, missä solmut ja särmät vuorottelevat on, polku (walk).

Solmu v_0 on polun W alkusolmu ja v_n loppusolmu. Voidaan sanoa, että W on polku solmusta v_0 solmuun v_n.

Määritelmä 3.24. Vrt. [4, s. 10], [6, luku 1.1.2] ja [28, s. 20]. Piiri on polku, jonka alku- ja loppusolmu on sama.

Määritelmä 3.25. Vrt. [6, luku 1.1.2] ja [28, s. 20]. Yksinkertainen polku (trail) on polku, jossa yksikään särmä ei esiinny useammin kuin kerran.

Määritelmä 3.26. Vrt. [4, s. 6–7], [6, luku 1.1.2] ja [28, s. 5]. Suora polku (path) on polku, jossa yksikään solmu ei esiinny useammin kuin kerran.

Lause 3.8. *Suora polku voidaan esittää yksinkertaisena graafina.*

Todistus. Olkoon P suora polku. Luodaan kaikista siihen kuuluvista solmuista joukko V ja kaikista siihen kuuluvista särmistä joukko E. Selvästi näin saadaan graafi $G = (V, E)$. Jos graafissa G olisi luuppi, niin kyseinen solmu toistuisi vähintään kahdesti polussa P, mikä on vastoin suoran polun määritelmää 3.26. Saman määritelmän nojalla graafissa G ei myöskään ole moninkertaisia särmiä. Täten suora polku P voidaan esittää yksinkertaisena graafina G. □

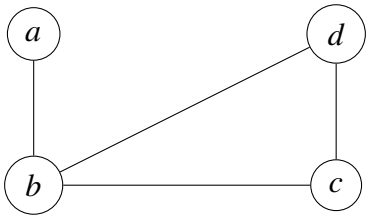
Määritelmä 3.27. Vrt. [4, s. 6–7], [6, luku 1.1.2] ja [28, s. 5]. Olkoon $P = v_0, v_0v_1, v_1, v_1v_2, \ldots, v_{n-1}v_n, v_n$ suora polku graafissa G. Jos $v_nv_0 \in G$, niin $C = v_0, v_0v_1, v_1, v_1v_2, \ldots, v_{n-1}v_n, v_n, v_nv_0$ on silmukka.

Lause 3.9. *Silmukka voidaan esittää yksinkertaisena graafina.*

Todistus. Vrt. lauseen 3.8 todistus. □

Määritelmä 3.28. Vrt. [4, s. 8], [6, luku 1.1.2] ja [28, s. 20]. Polun, piirin ja silmukan pituus, $|W|$, on siihen kuuluvien särmien lukumäärä.

Esimerkki 3.10. Vrt. [4, s. 7], [6, luku 1.1.2] ja [28, s. 6, 20–21]. Olkoon G alla esitetty graafi.



Kuva 3.6: Graafi G

Tällöin $W = b, bd, d, bd, b, ab, a$ on polku pituudeltaan 3, $Z = a, ab, b, ab, a$ on piiri pituudeltaan 2, $T = a, ab, b, bc, c, cd, d, bd, b$ on yksinkertainen polku pituudeltaan 4, $P = b, bc, c, cd, d$ on suora polku pituudeltaan 2 ja $C = b, bc, c, cd, d, bd$ on silmukka pituudeltaan 3. Alla on esitetty suora polku P yksinkertaisena graafina.

Kuva 3.7: Suora polku P

Myös graafi G on yksinkertainen, mutta sitä ei vastaa mikään polku, joten kaikkia yksinkertaisia graafeja ei voida esittää polkuina.

Polkujen avulla määritellään seuraavaksi solmujen väliset etäisyydet sekä graafin halkaisija ja sisäpiiri.

Määritelmä 3.29. Vrt. [4, s. 8] ja [28, s. 71]. Olkoot G graafi ja $a, b \in G$. Jos on olemassa polku solmusta a solmuun b, niin solmujen a ja b etäisyys, $dist_G(a, b)$, on lyhimmän niiden välisen suoran polun pituus. Muuten eli jos polkua ei ole, niin $dist_G(a, b) = \infty$.

Määritelmä 3.30. Vrt. [4, s. 8] ja [28, s. 71]. Olkoon G graafi. Graafin G halkaisija, $diam(G)$, on suurin etäisyys minkään sen kahden solmun välillä eli

$$diam(G) = \max_{a,b \in G} dist_G(a, b).$$

Määritelmä 3.31. Vrt. [4, s. 8]. Olkoon G graafi. Graafin G sisäpiiri, $girth(G)$, on lyhimmän graafissa G olevan silmukan pituus. Jos graafissa ei ole yhtään silmukkaa, niin $girth(G) = \infty$.

Esimerkki 3.11. Olkoon G esimerkissä 3.10 esitetty graafi G. Tällöin sen solmujen väliset etäisyydet ovat: $dist_G(a, b) = 1$, $dist_G(a, c) = dist_G(a, d) = 2$, $dist_G(b, c) = dist_G(b, d) = 1$ ja $dist(c, d) = 1$.

Koska kaikissa suuntaamattomissa graafeissa millä tahansa sen kahdella solmulla x ja y pätee $dist(x, y) = dist(y, x)$, niin graafin G halkaisija on $diam(G) = 2$.

Lisäksi $C = b, bc, c, cd, d, bd$ on ainoana silmukkana graafin G lyhin silmukka ja siten $girth(G) = 3$.

Vielä kulkemiseen liittyen osoitetaan intuitiivinen ajatus, että jos paikasta A päästään paikkaan B kulkemalla paikan C kautta useammin kuin kerran, niin paikasta A päästään paikkaan B myös kulkemalla paikan C kautta vain kertaalleen. Tämä selittää myös, miksi etäisyyden ja sisäpiirin määritelmissä on voitu käyttää suoraa polkua ja silmukkaa polun ja piirin sijaan.

Lause 3.10. *Vrt. [4, 10] ja [28, s. 21]. Olkoon W polku solmusta a solmuun b. Poistamalla polusta W ylimääräiset solmujen ja särmien toistot saadaan suora polku solmusta a solmuun b.*

Todistus. Olkoot W polku solmusta a solmuun b ja p polun pituus. Osoitetaan induktiolla polun pituuden suhteen, että poistamalla polusta W ylimääräiset solmujen ja särmien toistot saadaan suora polku solmusta a solmuun b.

Kun $p = 0$, niin W koostuu täsmälleen yhdestä solmusta eikä poistettavia ylimääräisiä solmujen ja särmien toistoja ole. Sen sijaan W on itsessään suora polku.

Tehdään induktio-oletus, että poistamalla polusta W' ylimääräiset solmujen ja särmien toistot saadaan suora polku solmusta a solmuun b, kun polun pituus on enintään $p - 1$.

Kun $p \geq 1$, niin W voi olla itsessään suora polku tai ei. Jos W ei ole suora polku, niin se sisältää vähintään yhden silmukan. Luodaan polku W', josta on poistettu kaikki silmukkaan kuuluvat solmut ja särmät alkusolmua lukuun ottamatta. Polun W' pituus on enintäään $p - 1$, joten induktio-oletuksen nojalla siitä saadaan tehtyä suora polku.

Tapausten $p = 0$ ja $p \geq 1$ sekä induktioperiaatteen nojalla poistamalla polusta W ylimääräiset solmujen ja särmien toistot saadaan suora polku solmusta a solmuun b. □

3.5 Graafien ominaisuuksia

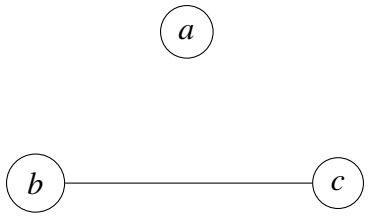
Tähän alalukuun on koottu erilaisia hyödyllisiä graafien ominaisuuksia. Aloitetaan määrittelemällä yhtenäisyys, joka mahdollistaa kulkemisen kaikkien solmujen välillä, ja komponentti, joka on tietynlainen aligraafi. Myös yhtenäisyyden määritelmässä on hyödynnetty lausetta 3.10 ja sen perusteella tarksteltu vain suoria polkuja.

Määritelmä 3.32. Vrt. [4, s. 10] ja [28, s. 21]. Olkoon G graafi. Jos kaikilla $a, b \in G$ on olemassa suora polku solmusta a solmuun b, niin G on yhtenäinen. Nollagraafi ei ole yhtenäinen.

Esimerkki 3.12. Esimerkin 3.10 molemmat graafit G ja P ovat yhtenäisiä, mutta esimerkin 3.13 graafi H ei ole.

Määritelmä 3.33. Vrt. [4, s. 11] ja [28, s. 22]. Olkoon G graafi ja H sen yhtenäinen aligraafi. Jos ei ole olemassa toista yhtenäistä graafin G aligraafia H', jonka aligraafi H on, niin H on graafin G komponentti.

Esimerkki 3.13. Olkoon H alla esitetty graafi.



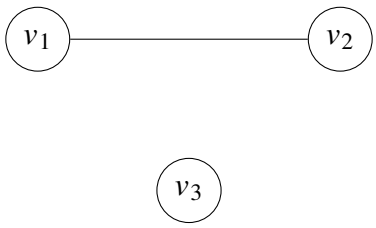
Kuva 3.8: Graafi H

Tällöin $(\{a\}, \emptyset)$ ja $(\{b, c\}, \{bc\})$ ovat sen komponentit.

Lopuksi määritellään isomorfia kahden graafin välillä. Isomorfiakuvauksen avulla saadaan rinnastettua rakenteeltaan samanlaiset graafit keskenään.

Määritelmä 3.34. Vrt. [4, s. 3] ja [28, s. 7]. Olkoot G ja H yksinkertaisia graafeja ja $f\colon V(G) \to V(H)$ bijektio, jolla $ab \in E(G)$ jos ja vain jos $f(a)f(b) \in E(H)$. Tällöin f on isomorfiakuvaus ja graafit G ja H ovat keskenään isomorfisia, jota merkitään $G \cong H$.

Esimerkki 3.14. Olkoot graafi H kuten esimerkissä 3.13 ja G alla esitetty graafi.



Kuva 3.9: Graafi G

Tällöin $H \cong G$ ja mahdolliset isomorfiakuvaukset ovat $f, g\colon V(H) \to V(G)$

$$f(u) = \begin{cases} v_3, & \text{kun } u = a, \\ v_1, & \text{kun } u = b \\ v_2, & \text{kun } u = c \end{cases} \text{ ja } g(u) = \begin{cases} v_3, & \text{kun } u = a, \\ v_2, & \text{kun } u = b \\ v_1, & \text{kun } u = c. \end{cases}$$

4 Laajenevien graafien perheet

Tässä luvussa tarkastellaan laajenevien graafien perheitä (expander graphs/families). Petit'n [17] mukaan nämä perheet muodostuvat jonosta graafeja, jotka ovat vahvasti yhtenäisiä, mutta sisältävät rajatusti särmiä. Ne ovat ikään kuin verkostoja, joissa huhut leviävät todella nopeasti. Kaikki tässä luvussa käsiteltävät graafit ovat suuntaamattomattomia ja saattavat sisältää moninkertaisia särmiä sekä luuppeja.

4.1 Laajenemissuhde ja sen merkitys

Ennen kuin voidaan määritellä laajenevien graafien perhe, niin täytyy tarkastella, miten saadaan varmistettua vahva yhtenäisyys ja mitä sillä oikeastaan edes tarkoitetaan. Tätä varten määritellään graafin solmujen osajoukolle raja ja todistetaan sen olevan sama kuin kyseisen osajoukon komplementin raja.

Määritelmä 4.1. Vrt. [10, s. 24] ja [8, s. 452]. Olkoot G graafi ja $S \subseteq V(G)$. Joukon S raja (boundary of S), ∂S, on joukko, johon kuuluvat ne graafin G särmät, joiden toinen päätepiste on joukossa S ja toinen joukossa $V(G) \setminus S$.

Lause 4.1. *Vrt. [10, s. 25]. Olkoon G graafi. Jokaisella $S \subseteq V(G)$ joukon S raja on sama kuin sen komplementin raja eli $\partial S = \partial V(G) \setminus S$.*

Todistus. Olkoon G graafi. Tällöin jokaisella $S \subseteq V(G)$ pätee $\partial S \subseteq \partial(V(G) \setminus S)$, sillä määritelmän 4.1 mukaan joukkoon ∂S kuuluvat särmät, joiden toinen päätepiste on joukossa S ja toinen joukossa $V(G) \setminus S$, ja jotka kaikki kuuluvat symmetrian nojalla myös joukkoon $\partial(V(G) \setminus S)$. Vastaavasti $\partial(V(G) \setminus S) \subseteq \partial S$. Täten $\partial S = \partial(V(G) \setminus S)$ eli joukon S raja on sama kuin sen komplementin raja jokaisella $S \subseteq V(G)$. □

Sitten määritellään laajenemissuhde. Tässä tarkasteltavien graafin solmujen osajoukkojen kokoa voidaan rajata edellisen lauseen perusteella, sillä joko osajoukon itsensä tai sen komplementin koko on aina vähintään puolet alkuperäisen joukon koosta.

Määritelmä 4.2. Vrt. [10, s. 25] ja [8, s. 452]. Olkoon G graafi. Graafin G laajenemissuhde (expansion ratio/isoperimetric constant) on

$$h(G) = \min \left\{ \frac{|\partial S|}{|S|} \right\},$$

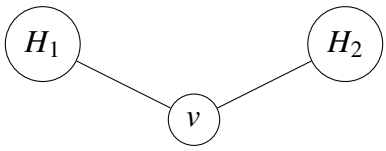
kun $S \subseteq V(G)$ ja $|S| \leq \frac{|V(G)|}{2}$.

Selvästi raja ja laajenemissuhde liittyvät yhtenäisyyteen. Nimittäin epäyhtenäisellä graafilla on olemassa solmujen osajoukko, jonka raja on tyhjä joukko ja laajenemissuhde siten nolla. Seuraavien esimerkkien nojalla näyttäisikin siltä, että mitä vahvemmin yhtenäinen tai lähempänä täydellistä graafia kyseinen graafi on sitä suurempi sen laajenemissuhde on.

Esimerkki 4.1. Olkoot H täydellinen graafi, jolle $V(H) = \{a, b, c, d, e\}$, ja $S_0 = \{a, b\}$ sekä $S_1 = \{a\}$. Tällöin $\partial S_0 = \{ac, ad, ae, bc, bd, be\}$ ja $\partial S_1 = \{ab, ac, ad, ae\}$. Ensinnäkin koska $|V(H)|/2 = 2,5$, niin laajenemissuhdetta laskiessa ei tarvitse määritelmän 4.2 mukaan huomioida osajoukkoja, joissa on enemmän kuin kaksi solmua. Lisäksi koska H on täydellinen graafi, niin kaikilla $S_2 \subseteq V(H)$, joilla $|S_2| = 2$, pätee $|\partial S_2| = |\partial S_0|$. Vastaavasti kaikilla $S_3 \subseteq V(H)$, joilla $|S_3| = 1$, pätee $|\partial S_3| = |\partial S_1|$. Täten voidaan laajenemissuhde laskea seuraavasti

$$h(H) = \min\left\{\frac{|\partial S_0|}{|S_0|}, \frac{|\partial S_1|}{|S_1|}\right\} = \min\left\{\frac{6}{2}, \frac{4}{1}\right\} = 3.$$

Esimerkki 4.2. Olkoot H_1 ja H_2 kuten edellisen esimerkin graafi H ja G alla esitetty graafi, jossa siis $V(G) = V(H_1) \cup V(H_2) \cup \{v\}$ ja $E(G) = E(H_1) \cup E(H_2) \cup \{h_1v, h_2v\}$, missä $h_1 \in H_1$ ja $h_2 \in H_2$. Koska H_1 ja H_2 ovat täydellisiä graafeja, niin ei ole laajenemissuhteen kannalta merkitystä, mitkä niiden solmuista on solmun v vierussolmut.



Kuva 4.1: Graafi G

Ensinnäkin koska $|V(G)|/2 = 5,5$, niin laajenemissuhdetta laskiessa ei tarvitse määritelmän 4.2 mukaan huomioida osajoukkoja, joissa on enemmän kuin viisi solmua. Lisäksi koska graafi G on yhtenäinen, niin pienin mahdollinen laajenemissuhde on $1/5$, sillä jokainen joukko solmuja on yhteydessä muihin solmuihin vähintään yhdellä särmällä. Tässä tapauksessa pienin mahdollinen laajenemissuhde saadaan osajoukoilla $V(H_1) \subseteq V(G)$ ja $V(H_2) \subseteq V(G)$. Täten

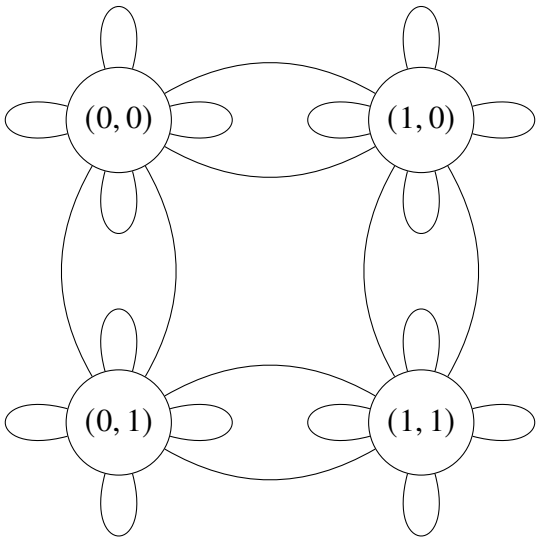
$$h(G) = \frac{|\partial V(H_1)|}{|V(H_1)|} = \frac{|\partial V(H_2)|}{|V(H_2)|} = \frac{1}{5}.$$

Seuraavaksi määritellään nollasta pois sitominen. Tämän avulla saadaan myös määriteltyä laajenevien graafien perhe, jossa särmien määrää rajoitetaan asettamalla se säännölliseksi.

Määritelmä 4.3. Vrt. [10, s. 27]. Olkoon $(a_n) = a_0, a_1, \ldots, a_n \in \mathbb{R}$ jono, missä $n \geq 0$ on kokonaisluku. Kun on olemassa reaaliluku $\epsilon > 0$, jolla kaikilla kokonaisluvuilla $i \in [0, n]$ pätee $a_i \geq \epsilon$, niin jono (a_n) on sidottu pois nollasta (bounded away from zero).

Määritelmä 4.4. Vrt. [10, s. 27] ja [8, s. 452]. Olkoon (G_n) jono d-säännöllisiä graafeja, joilla $|V(G_n)| \to \infty$, kun $n \to \infty$. Kun laajenemissuhteiden jono $(h(G_n))$ on sidottu pois nollasta, niin (G_n) on laajenevien graafien perhe (expander graphs/family). Yksittäistä laajenevien graafien perheen graafia voidaan kutsua laajenevaksi graafiksi.

Esimerkki 4.3. Vrt. [8, s. 453 ja 503]. Olkoon G_n graafi, jossa n on positiivinen kokonaisluku ja jolle $V(G_n) = \mathbb{Z}_n \times \mathbb{Z}_n$. Kaikilla $x, y \in \mathbb{Z}_n$ solmun (x, y) vierussolmut ovat $(x+y, y)$, $(x-y, y)$, $(x, y+x)$, $(x, y-x)$, $(x+y+1, y)$, $(x-y+1, y)$, $(x, y+x+1)$ ja $(x, y-x+1)$, missä kaikki laskutoimitukset on suoritettu $\mod n$. Tämä määrittelee graafin G_n särmät ja tekee siitä 8-säännöllisen. Alla on esitetty graafi G_2.



Kuva 4.2: Laajeneva graafi G_2

Hoory, Linial ja Widgerson todistavat artikkelissaan [8, s. 503- 508], että tällöin (G_n) on laajenevien graafien perhe.

Yllä olevassa esimerkissä esitelty laajenevien graafien perhe tunnetaan Margulis'n perheenä (Margulis construction). Hoory, Linial ja Widgerson [8] kertovat sen olevan vanhin tarkka kuvaus mistään laajenevien graafien perheestä.

4.2 Spektrin aukon koon merkitys

Myöhemmin tässä alaluvussa todistetaan, että laajenevien graafien perheet voitaisin yhtä hyvin määritellä spektrien aukkojen kokoishin perustuen. Sitä varten tarvitaan kuitenkin useita spektraaligraafiteoriasta nousevia tuloksia. Niiden käsittelemiseksi määritellään ensin tavallinen pistetulo ja normi. Sitten tarkastellaan tiettyä pistetuloa graafin solmuja käsittelevien funktioiden välillä.

Määritelmä 4.5. Vrt. [10, s. 10]. Olkoot $f, g \in F(S, \mathbb{C})$, missä S on jokin äärellinen joukko. Funktioiden f ja g välinen pistetulo on

$$\langle f, g \rangle = \sum_{s \in S} f(s)\overline{g(s)}$$

ja funktion f normi on

$$||f|| = \sqrt{\langle f, f \rangle}.$$

Lause 4.2. *Vrt. [10, s. 23-24]. Olkoot $G = (V, E)$ d-säännöllinen graafi ja $f \in F(V, \mathbb{C})$. Suunnitaan graafi G. Tällöin saadaan pistetulo*

$$\langle \Delta f, f \rangle = \sum_{\overline{e} \in E} |f(e^+) - f(e^-)|^2.$$

Todistus. Olkoot $G = (V, E)$ d-säännöllinen graafi, $f \in F(V, \mathbb{C})$ ja $g \in F(E, \mathbb{C})$. Tarkastellaan ensin määritelmien 3.22 ja 4.5 avulla pistetuloa

$$\begin{aligned}
\langle \delta f, g \rangle &= \sum_{\overline{e} \in E} (\delta f)(\overline{e}) \overline{g(\overline{e})} \\
&= \sum_{\overline{e} \in E} [f(e^+) - f(e^-)] \overline{g(\overline{e})} \\
&= \sum_{\overline{e} \in E} f(e^+) \overline{g(\overline{e})} - \sum_{\overline{e} \in E} f(e^-) \overline{g(\overline{e})} \\
&= \sum_{v \in V} f(v) \sum_{\overline{e} \in E, v = e^+} \overline{g(\overline{e})} - \sum_{v \in V} f(v) \sum_{\overline{e} \in E, v = e^-} \overline{g(\overline{e})} \\
&= \sum_{v \in V} f(v) \overline{(\delta^* g)(v)} \\
&= \langle f, \delta^* g \rangle.
\end{aligned}$$

Tämän perusteella saadaan

$$\langle \Delta f, f \rangle = \langle \delta^* \delta f, f \rangle = \overline{\langle f, \delta^* \delta f \rangle} = \overline{\langle \delta f, \delta f \rangle} = \langle \delta f, \delta f \rangle$$

ja edelleen

$$\langle \delta f, \delta f \rangle = \sum_{\overline{e} \in E} (f(e^+) - f(e^-)) \overline{(f(e^+) - f(e^-))} = \sum_{\overline{e} \in E} |f(e^+) - f(e^-)|^2.$$

□

Jatkossa merkitään $F(S, \mathbb{R})$ joukolle, joka sisältää kaikki funktiot $f \colon S \to \mathbb{R}$, missä S on jokin äärellinen joukko. Lisäksi merkitään $F_0(S, \mathbb{R})$ joukolle, joka sisältää ne funktiot $f \colon S \to \mathbb{R}$, joilla $\sum_{s \in S} f(s) = 0$. Näiden merkintöjen avulla voidaan yksinkertaistaa seuraavan lauseen esitystä. Se tunnetaan Rayleigh-Ritzin lauseeena ja se osoittaa yhteyden toiseksi suurimman ominaisarvon ja tietyn pistetulon välillä. Erityisesti sen seurauksena voidaan osoittaa spektrin aukon koon yhteys edellisessä lauseessa tarkasteltuun pistetuloon.

Lause 4.3. *Vrt. [10, s. 30-31]. Olkoon $G = (V, E)$ d-säännöllinen graafi. Tällöin*

$$\lambda_1(G) = \max_{f \in F_0(V, \mathbb{R}), ||f||=1} \langle Af, f \rangle.$$

Todistus. Ks. [10, s. 30-31]. □

Seuraus 4.4. *Vrt. [10, s. 30-31]. Olkoon $G = (V, E)$ d-säännöllinen graafi. Tällöin*

$$d - \lambda_1(G) = \min_{f \in F_0(V, \mathbb{R}), ||f||=1} \langle \Delta f, f \rangle.$$

Todistus. Olkoot $G = (V, E)$ d-säännöllinen graafi ja A sen vierusmatriisi. Lauseen 4.3 mukaan

$$\lambda_1(G) = \max_{f \in F_0(V,\mathbb{R}), ||f||=1} \langle Af, f \rangle$$

ja lauseen 3.7 mukaan $\Delta = dI - A$ eli $A = dI - \Delta$. Nämä yhdistämällä saadaan

$$\lambda_1(G) = \max_{f \in F_0(V,\mathbb{R}), ||f||=1} \langle (dI - \Delta)f, f \rangle.$$

Tarkastellaan määritelmän 4.5 avulla tarkemmin pistetuloa

$$\begin{aligned}
\langle (dI - \Delta)f, f \rangle &= \sum_{\overline{e} \in E} ((dI - \Delta)f)(\overline{e})\overline{f(\overline{e})} \\
&= \sum_{\overline{e} \in E} \Big((dIf)(\overline{e}) - (\Delta f)(\overline{e}) \Big) \overline{f(\overline{e})} \\
&= d \sum_{\overline{e} \in E} f(\overline{e})\overline{f(\overline{e})} - \sum_{\overline{e} \in E} (\Delta f)(\overline{e}\overline{f(\overline{e})} \\
&= d - \langle \Delta f, f \rangle.
\end{aligned}$$

Tällöin

$$\lambda_1(G) = \max_{f \in F_0(V,\mathbb{R}), ||f||=1} \langle Af, f \rangle = \max_{f \in F_0(V,\mathbb{R}), ||f||=1} (d - \langle \Delta f, f \rangle),$$

mistä saadaan edelleen

$$\lambda_1(G) = d - \min_{f \in F_0(V,\mathbb{R}), ||f||=1} \langle \Delta f, f \rangle.$$

Näin ollen

$$d - \lambda_1(G) = \min_{f \in F_0(V,\mathbb{R}), ||f||=1} \langle \Delta f, f \rangle.$$

□

Seuraavassa lauseessa voidaan aiemmin esitettyjen tulosten avulla osoittaa laajenemissuhteen ja spektrin aukon koon välinen yhteys. Tämän jälkeen oleva lause on käytännössä vaihtoehtoinen määritelmä laajenevien graafien perheille perustuen spektrien aukkojen kokoihin.

Lause 4.5. *Vrt. [10, s. 31-32] ja [8, s. 454 ja 474-475]. Olkoon G d-säännöllinen graafi. Tällöin*

$$h(G) \geq \frac{d - \lambda_1(G)}{2}.$$

Todistus. Olkoon $G = (V, E)$ d-säännölinen graafi. Määritelmän 4.2 nojalla voidaan valita $S \subseteq V$, jolla $|S| \leq |V|/2$ ja $h(G) = |\partial S|/|S|$. Merkitään $a = |V \setminus S|$ ja $b = |S|$. Määritellään funktiot $f, g \in F(V, \mathbb{R})$ seuraavasti

$$g(v) = \begin{cases} a & \text{, kun } v \in S \\ -b & \text{, kun } v \in V \setminus S \end{cases} \text{ ja } f(v) = \frac{g(v)}{||g||}.$$

Itseasiassa $f, g \in F_0(V, \mathbb{R})$, sillä

$$\sum_{v \in V} g(v) = \sum_{v \in S} a - \sum_{v \in V \setminus S} b = ba - ab = 0.$$

Suunnitaan graafi G. Tällöin lauseen 4.2 mukaan

$$\langle \Delta g, g \rangle = \sum_{\overline{e} \in E} |g(e^+) - g(e^-)|^2.$$

Koska kaikilla särmillä, joiden molemmat päätepisteet ovat joko joukossa S tai joukossa $V \setminus S$ summattava neliö menee nollaksi funktion g määritelmän antaessa solmuille e^+ ja e^- saman arvon, niin saadaan edelleen

$$\sum_{\overline{e} \in E} |g(e^+) - g(e^-)|^2 = \sum_{\overline{e} \in \partial S} (a+b)^2 = |\partial S|(a+b)^2.$$

Lisäksi

$$||g||^2 = \langle g, g \rangle = \sum_{v \in V} (g(v))^2 = \sum_{v \in S} a^2 + \sum_{v \in V \setminus S} (-b)^2 = ba^2 + ab^2 = ab(a+b).$$

Näin ollen

$$\langle \Delta f, f \rangle = \left\langle \Delta \frac{g}{||g||}, \frac{g}{||g||} \right\rangle = \frac{\langle \Delta g, g \rangle}{||g||^2} = \frac{|\partial S|(a+b)^2}{ab(a+b)} = \left(1 + \frac{b}{a}\right) \frac{|\partial S|}{|S|} = \left(1 + \frac{b}{a}\right) h(G).$$

Koska $b \leq |V|/2$, niin $a \geq b$ ja $0 \leq b/a \leq 1$, niin voidaan arvioida

$$\langle \Delta f, f \rangle = \left(1 + \frac{b}{a}\right) h(G) \leq 2h(G).$$

Seurauksen 4.4 nojalla tiedetään, että $d - \lambda_1(G) \leq \langle \Delta f, f \rangle$ ja näin ollen

$$d - \lambda_1(G) \leq \langle \Delta f, f \rangle \leq 2h(G),$$

mistä saadaan edelleen

$$h(G) \geq \frac{d - \lambda_1(G)}{2}.$$

□

Lause 4.6. *Vrt. [10, s. 32] ja [8, s.454]. Olkoon (G_n) jono d-säännöllisiä graafeja, joilla $|V(G_n)| \to \infty$, kun $n \to \infty$. Kun spektrien aukkojen ko'oista muodostettu jono $(d - \lambda_1(G_n))$ on sidottu pois nollasta, niin (G_n) on laajenevien graafien perhe.*

Todistus. Olkoon (G_n) jono d-säännöllisiä graafeja, joilla $|G_n| \to \infty$, kun $n \to \infty$. Kun jono $(d - \lambda_1(G_n))$ on sidottu pois nollasta, niin määritelmän 4.3 mukaan on olemassa reaaliluku $\epsilon > 0$, jolla kaikilla n pätee $d - \lambda_1(G_n) \geq \epsilon$. Yhdessä lauseen 4.5 kanssa saadaan

$$h(G_n) \geq \frac{d - \lambda_1(G_n)}{2} \geq \frac{\epsilon}{2}.$$

Täten on löydetty reaaliluku $\epsilon' = \frac{\epsilon}{2} > 0$, jolla kaikilla n pätee $h(G_n) \geq \epsilon'$, joten laajenemissuhteiden jono $(h(G_n))$ on sidottu pois nollasta. Näin ollen määritelmän 4.4 mukaan (G_n) on laajenevien graafien perhe. □

Tämän alaluvun lopuksi määritellään Ramanujanin graafien perheet. Ne ovat laajenevien graafien perheitä, jotka ovat jo määritelmänsä perusteella yhteydessä graafien ominaisarvoihin.

Määritelmä 4.6. Vrt. [10, s. 68-69] ja [17, s. 78-79]. Olkoon (G_n) jono d-säännöllisiä graafeja. Kun $\lambda(G_n) \leq 2\sqrt{d-1}$, missä $\lambda(G_n) = \max\{|\lambda_1|, |\lambda_{n-1}|\}$, niin jono (G_n) muodostaa perheen Ramanujanin graafeja.

Lause 4.7. *Vrt. [10, s.70]. Olkoon (G_n) perhe d-säännöllisiä Ramanujanin graafeja. Jos $d \geq 3$, niin (G_n) on laajenevien graafien perhe.*

Todistus. Olkoon (G_n) perhe d-säännöllisiä Ramanujanin graafeja. Määritelmän 4.6 mukaan $\lambda(G_n) \leq 2\sqrt{d-1}$, missä $\lambda(G_n) = \max\{|\lambda_1|, |\lambda_{n-1}|\}$. Tällöin $\lambda_1(G_n) \leq 2\sqrt{d-1}$. Tämän nojalla $d - \lambda_1(G_n) \geq d - 2\sqrt{d-1}$. Kun $d \geq 3$, niin $d - \lambda_1(G_n) \geq 3 - 2\sqrt{3-1} \approx 0,17$. Voidaan siis valita $\epsilon = 0,1$, jolla jono $(d - \lambda_1(G_n))$ on sidottu pois nollasta ja siten lauseen 4.6 mukaan (G_n) on laajenevien graafien perhe. □

5 Laajenevia hashfunktioita

Tässä luvussa esitellään laajenevien graafien perheisiin perustuvia hashfunktioita, jotka tunnetaan laajenevina hashfunktioina (expander hash). Niissä viestin perusteella luodaan polku laajenevassa graafissa ja hasharvo saadaan polun viimeisestä solmusta. Pohjalla oleva graafiteoria tekee tälläisistä hashfunktioista todistettavissa olevia, mikä mahdollistaa tietoturvallisuuden tutkimisen eli kryptoanalyysin irrallaan itse hashfunktion rakenteesta. Laajenevien hashfunktioiden teoria levittyykin laajalti matematiikan eri osa-alueille sekä tietotekniikkaa, minkä johdosta yksityiskohtainen tarkastelu vaatisi esitietoja kaikilta osa-alueilta. Tässä tutkielmassa esitellään idea ja annetaan tarpeeksi yksityiskohtia kokonaiskuvan hahmottamiseksi. Lisäksi viitataan mahdollisimman täsmällisesti eteenpäin useisiin lähteisiin loppujen esitietojen ja yksityiskohtien löytämiseksi.

5.1 Laajenevat hashfunktiot ja niiden tietoturvaominaisuudet

Satunnaispolku (random walk) on polku, jonka solmut ja särmät on valittu satunnaisesti. Petit'n [17, s. 94] mukaan laajenevien graafien satunnaispolkuja koskevat tulokset voidaan yleistää koskemaan myös laajenevia hashfunktioita. Siksi aluksi määritelläänkin satunnaispolut ja tarkastellaan niitä laajenevissa graafeissa.

Määritelmä 5.1. Vrt. [8, s. 458] . Olkoon $G = (V, E)$ d-säännöllinen graafi. Satunnaispolku (random walk) on polku, joka alkaa normaalijaukaman avulla valitusta solmusta $v_0 \in V$ ja jonka jokainen särmä, ja siten myös sitä seuraava solmu, valitaan satunnaisesti niiden särmien joukosta, joilla on päätepiste kyseisessä solmussa. Satunnaispolku ilman takaisin paluuta muodostetaan muuten samoin, mutta samaa särmää ei saa valita kahdesti peräkkäin.

Lause 5.1. *Vrt. [17, s. 79]. Olkoon G laajeneva graafi. Tällöin graafissa G satunnaispolkuun kuuluvat solmut muodostavat joukon, joka on hyvin lähellä satunnaisesti valittua osajoukkoa kaikkien solmujen joukosta $V(G)$. Lisäksi eri satunnaispolkujen viimeisten solmujen jakauma lähestyy hyvin nopeasti normaalijakaumaa.*

Lisää tietoa satunnaispoluista laajenevissa graafeissa löytyy esimerkikisi Petit'n väitöskirjasta [17] sivuilta 79-84 ja niiden yhteydestä laajeneviin hashfunktioihin sivuilta 94-95. Hoory, Linial ja Widgerson puolestaan käsittelevät satunnaispolkuja laajenevissa graafeissa artikkelissaan [8] sivuilla 458-469. He esittävät lisäksi useita muita käytännön sovelluksia niille. Toisinaan satunnaispolut esitetään myös Markov'n ketjuina, kuten Maricq on tehnyt opinnäytetyössään [15, s. 41-43]. Tämän pohjalta vaikuttaakin lupaavalta yleisesti hyödyntää satunnaispolkuja laajenevissa hashfunktioissa.

Määritelmä 5.2. Vrt. [2, s. 96-97] ja [17, s. 86-89]. Olkoon (G_n) d-säännöllinen laajenevien graafien perhe. Laajeneva hashfunktio (expander hash) on seuraavanlainen hashfunktio $h\colon M \to \mathbb{B}_s$. Olkoon $m \in M$ binäärilukuesitys viestistä, jonka hasharvo halutaan laskea.

1. Jaetaan viesti m lohkoihin $m_0, \ldots, m_{j-1}$, missä jokaisella $i \in [0, j-1]$ pätee $m_i \in [0, d-2]$.

2. Valitaan graafi G perheestä (G_n) ottaen huomioon, että sen kaikki solmut voidaan nimetä yksikäsitteisesti s-bittisillä kokonaisluvuilla, eli $|G| \leq 2^s$. Valitaan jokin solmu graafista G ja alustetaan se apumuuttujaan V.

3. Muodostetaan graafissa G ilman takaisin paluuta polku, joka alkaa apumuuttujan V määräämästä solmusta ja missä jokaista lohkoa m_i käytetään vuorollaan valitsemaan seuraava särmä. Päivitetään apumuuttujaan V aina solmu, johon on päästy.

4. Kun kaikki lohkot on käsitelty, niin viestin m hasharvo saadaan apumuuttujaan V viimeisenä tallennetun solmun nimestä.

Avaimellinen laajeneva hashfunktio on laajeneva hashfunktio, jonka salausavain k sisältää tiedon apumuuttujaan V kohdassa 2 ensimmäiseksi valittavasta solmusta.

Laajenevan hashfunktion yleinen määrittely on esitetty yksityiskohtaisemmin Petit'n väitöskirjassa [17] sivuilla 86-89. Nyt tarkastellaan, millaisina ongelmina hashfunktion tietoturvaominaisuudet laajenevissa hashfunktioissa esiintyvät.

Lause 5.2. *Vrt. [17, s. 90-91]. Olkoot $l > 1$ kokonaisluku, $h\colon M \to \mathbb{B}_n$ laajeneva hashfunktio, G siihen liittyvä laajeneva graafi ja v_0 polkujen aloitussolmu. Tällöin hashfunktion h tietoturvaominaisuudet vastaavat seuraavia ongelmia.*

- ***Polkuongelma*** *(path problem): Olkoon $v \in V(G)$. Jos on vaikeaa löytää polku W solmusta v_0 solmuun v, missä $|W| \leq l$, niin h on alkukuvaresistentti.*

- ***Kahden polun ongelma*** *(two-path problem): Jos on vaikeaa löytää polut W ja W', joilla $|W|, |W'| \leq l$ ja niillä on samat alku- ja loppusolmut , niin h on toisen alkukuvan suhteen resistentti.*

- ***Luuppiongelma*** *(cycle problem): Jos on vaikeaa löytää luuppi C, jolla $|C| \leq l$, niin h on toisen alkukuvan suhteen resistentti.*

- ***Rajoitettu kahden polun ongelma*** *(constrained two-paths problem): Jos on vaikeaa löytää polut W ja W' solmusta v_0 johonkin solmuun $v \in V(G)$, missä $|W| \leq l$ ja $W \neq W'$, niin h on törmäysresistentti.*

- ***Rajoitettu luuppiongelma*** *(constrained cycle-problem): Jos on vaikeaa löytää luuppi C, jolla $|W| \leq l$ ja joka kulkee solmun v_0 kautta, niin h on törmäysresistentti.*

Näissä ongelmissa esiintyvä polkujen pituutta rajoittava kokonaisluku l on valittavissa tilanteen mukaan; mitä lyhyempien polkujen löytäminen on vaikeaa, niin sitä lyhyemmille viesteille kyseinen hashfunktio on turvallinen. Lisäksi todella pitkien polkujen laskemisessa voi kestää tietokoneella niin kauan, että ei enää ole tietoturvan

kannalta oleellista, kuinka helppoja laskut ovat, sillä aikavaatimus itsessään takaa turvallisuuden vähintään niin pitkäksi ajaksi.

Lisäksi on olemassa myös muita vain laajenevia hashfunktioita koskevia tietoturvaominaisuuksia, jotka liittyvät erityisesti niissä käytettävien laajenevien graafien ominaisuuksiin. Kutsutaan niitä siis laajenevan graafin tietoturvaominaisuuksiksi. Ne liittyvät graafin toiseksi suurimpaan ominaisarvoon, halkaisijaan, luuppien pituuksiin sekä vierussolmujen laskennan tehokkuuteen. Tärkeää on huomioida, että määritelmän 5.2 kohdassa 2 valitulla graafilla on kyseiset tietoturvaominaisuudet.

Määritelmä 5.3. Vrt. [17, s. 99]. Olkoot $h\colon M \to \mathbb{B}_n$ laajeneva hashfunktio ja G siihen liittyvä laajeneva graafi. Tällöin graafin G tietoturvaominaisuuksiin viitataan seuraavin termein.

1. **Nopeasti laajenevuus** (large expansion): Jos graafin G toiseksi suurin ominaisarvo on suuri, niin graafi G on nopeasti laajeneva.

2. **Pieni halkaisija** (short diameter): Jos graafin G halkaisija on pieni, niin graafi G on pieni halkaisijaltaan.

3. **Suuri sisäpiiri** (large girth): Jos graafin G sisäpiiri on suuri, niin graafi G on suuri piiriltään.

4. **Tehokkuus** (efficiency): Jos graafin G kaikilla solmuilla saadaan vierussolmut laskettua tehokkaasti, niin graafi G on tehokas.

Aiempien tietoturvaominaisuuksien tapaan suuruus, pienuus ja tehokkuus ovat tilanneriippuvaisia. Petit'n [17, s. 99] mukaan nopeasti laajenevuus varmistaa, että lyhyiden viestien hasharvot jakautuvat tasaisesti arvojoukkoon, ja pieni halkaisija sen, että jokainen solmu on jonkin lyhyen viestin hasharvo. Suuri piiri puolestaan varmistaa, että viestit, joilla syntyy törmäys, eroavat tarpeeksi toisistaan. Vielä tehokkuus antaa turvaa sivukanavahyökkäyksiä (side-channel attacks) vastaan, joissa hasharvon laskentaan käytetystä ajasta, virrankulutuksesta tai muusta varsinaisesti hashfunktion määrittelyyn riippumattomasta seikasta pyritään päättelemään jotakin itse viestistä.

5.2 Esimerkkejä laajenevista hashfunktioista

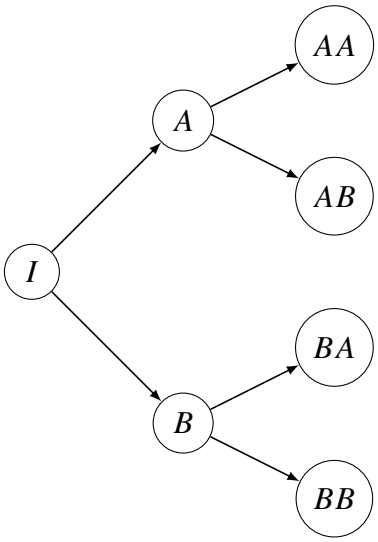
Tässä alaluvussa esitellään kaiken kaikkiaan kolmen laajenevan hashfunktion idea sekä annetaan lisätiedoissa viitteet myös viiteen muuhun. Esiteltävät hashfunktiot on valittu niin, että ne perustuvat kaikki eri laajenevien graafien perheisiin ja ovat tietoturvaominaisuuksiltaan erilaisia. Kuitenkaan niiden yksityiskohtainen määrittely todistuksineen ei ole mielekästä kokonaiskuvan hahmottamisen kannalta.

Ranskalainen Zémor oli ensimmäinen, joka julkaisi ehdotuksen laajenevasta hashfunktiosta vuonna 1991 [32]. Yhdessä Tillichin kanssa he mursivat kyseisen hashfunktion törmäysresistanssin alkuperäisillä parametreilla vuonna 1994 julkaistussa artikkelissaan [26] esittäen samalla uusia parempia parametreja. Samana vuonna he kuitenkin julkaisivat toisenkin samaan ideaan perustuvan ehdotuksen [27].

Esimerkki 5.1. Vrt. [17, s. 101] ja [27, s. 41]. Olkoot x muuttuja joukossa $\{0, 1\}$ ja

$$I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, A = \begin{bmatrix} x & 1 \\ 1 & 0 \end{bmatrix} \text{ sekä } B = \begin{bmatrix} x & x+1 \\ 1 & 1 \end{bmatrix}.$$

Laajenevien graafien perhe G_n muodostuu graafeista, jonka solmuina ovat binäärimatriisien I, A ja B tulona saatavat matriisit jollakin jaottomalla polynomilla rajoitettuna. Nämä tulot määrittävät myös särmät alla olevan kuvan mukaisesti.



Kuva 5.1: Osa perheen G_n laajenevasta graafista

Zémor-Tillichin laajeneva hashfunktio on $ZT\colon M \to \mathbb{B}_s$ ja määritellään seuraavasti. Olkoon $m \in M$ binäärilukuesitys viestistä, jonka hasharvo halutaan laskea.

1. Jaetaan viesti m yhden bitin lohkoihin $m_0, \ldots, m_{j-1}$.

2. Valitaan laajeneva graafi G perheestä (G_n) ottaen huomioon, että sen kaikki solmut voidaan nimetä yksikäsitteisesti s-bittisillä kokonaisluvuilla, eli $|G| \leq 2^s$. Valitaan solmu I ja alustetaan se apumuuttujaan V.

3. Jokaisella kokonaisluvulla $i \in [0, j-1]$ päivitetään apumuuttuja V seuraavalla tavalla
$$V = \begin{cases} VA & \text{, kun } m_i = 0 \\ VB & \text{, kun } m_i = 1. \end{cases}$$
Näin muodostetaan graafissa G polku, joka alkaa apumuuttujan V määräämästä solmusta ja missä jokaista lohkoa m_i käytetään vuorollaan valitsemaan seuraava särmä.

4. Kun kaikki lohkot on käsitelty, niin viestin m hasharvo saadaan apumuuttujaan V viimeisenä tallennetun solmun nimestä.

ZT-hashfunktiosta on lisää tietoa Petit'n väitöskirjassa [17] sivuilla 101-102 sekä Tillichin ja Zémorin alkuperäisessä julkaisussa [27]. Lisäksi Petit'n väitöskirjan sivuilla 111-146 on laaja kryptoanalyysi tämän hashfunktion tietoturvaominaisuuksista ja muutamia uusia muunnoksia siitä. Kaiken kaikkiaan hän esittää ZT-hashfunktion säilyttäneen törmäysresistanssinsa, kunhan sen parametrit ovat tarpeeksi suuria ja huolella valittuja. Kuitenkin vuonna 2011 Grassl, Ilić, Magliveras ja Steinwandt onnistuivat artikkelissaan [5] lopulta murtamaan niin alkuperäisen ZT-hashfunktion kuin Petit'n muunnostenkin törmäysresistanssin, mutta yksisuuntaisuus on edelleen voimassa.

Vuonna 2007 julkaisivat Charles, Lauter ja Goren oman ehdotuksensa laajenevasta hashfunktiosta [2], joka syntyi yhdysvaltalais-kanadalaisena yhteistyönä. Se perustuu vuonna 1988 Lubotzkyn, Phillipsin ja Sarnakin löytämiin laajeneviin graafeihin [12] ja se tunnetaankin LPS-hashfunktiona.

Esimerkki 5.2. Vrt. [2, s. 106-107] ja [17, s. 102-103]. Olkoot l ja p keskenään eri suuret alkuluvut, joista l on pieni ja p on suuri, jotka molemmat ovat 1 modulo 4 ja lisäksi l on neliönjäännös (quadratic residue) modulo p. Laajenevien graafien perhe G_n muodostuu eri tavoista valita nämä alkuluvut ja yksittäistä kyseisen perheen graafia merkitään $X_{l,p}$. Graafin $X_{l,p}$ solmut ovat ne kääntyvät $2x2$-matriisit, joiden alkiot ovat $\mathbb{Z}_p$-kertoimisia funktioita, joiden determinantti on 1, ja kaikilla $A \in X_{l,p}$ pätee $A = -A$.

Graafin $X_{l,p}$ särmät saadaan matriisien tuloilla seuraavalla tavalla. Olkoon $k \geq 0$ kokonaisluku. Merkitään $(a_0, a_1, a_2, a_3)_k$ yhtälön

$$g_0^2 + g_1^2 + g_2^2 + g_3^2 = l$$

kokonaislukuratkaisuja, missä muuttujan g_0 arvo a_0 on positiivinen ja pariton ja muiden muuttujien g_1, g_2 ja g_3 arvot a_1, a_2 ja a_3 ovat parillisia. Tälläisiä keskenään eri suuria ratkaisuja on yhteensä $l+1$ kappaletta. Olkoon lisäksi i kokonaisluku, jolla $i^2 \equiv -1 \mod p$. Joukko $\mathcal{G}$ muodostuu matriiseista $G_0, \ldots, G_l$, jotka muodostetaan äskeisen yhtälön ratkaisusta seuraavalla tavalla jokaisella $k \in [0, l]$ käyttäen eri ratkaisua $(a_0, a_1, a_2, a_3)_k$

$$G_k = \begin{bmatrix} a_0 - ia_1 & a_2 - ia_3 \\ -a_2 - ia_3 & a_0 + ia_1 \end{bmatrix}.$$

Solmun A vierussolmut ovat solmut GA kaikilla $G \in \mathcal{G}$. Näin ollen graafi $X_{l,p}$ on $l+1$-säännöllinen.

Näihin Lubotzky-Phillips-Sarnakin laajeneviin graafeihin perustuva hashfunktio on $LPS\colon M \to \mathbb{B}_s$. Usein valitaan $l = 5$ ja p on jokin 1024-bittinen aiemmat ehdot täyttävä alkuluku. Tällöin LPS-hashfunktio määritellään seuraavasti. Olkoon $m \in M$ binäärilukuesitys viestistä, jonka hasharvo halutaan laskea.

1. Jaetaan viesti m lohkoihin $m_0, \ldots, m_{j-1}$, missä jokaisella $k \in [0, j-1]$ pätee $m_k \in \{00, 01, 10, 11, 100\}$ ja täten $m_k \in [0, 4]$.

2. Valitaan graafi $X_{5,p}$ perheestä (G_n) ottaen huomioon, että sen kaikki solmut voidaan nimetä yksikäsitteisesti s-bittisillä kokonaisluvuilla, eli $|G| \leq 2^s$. Valitaan jokin solmu graafista $X_{5,p}$ ja alustetaan se apumuuttujaan V. Alustetaan myös jokin solmun V vierussolmuista apumuuttujaan V_{-1}.

3. Jokaisella k, jolla $k \in [0, j-1]$, valitaan lohkon m_k määrittämä kerroinmatriisi G joukosta $\mathcal{G}^*$ ja päivitetään apumuuttujiin $V_{-1} = V$ ja $V = GV$. Tässä $\mathcal{G}^* = \mathcal{G} \setminus G'$, missä matriisilla G' saadaan $GV = V_{-1}$ eli aiheutettaisiin takaisin paluu edelliseen solmuun.

   Näin muodostetaan graafissa $X_{5,p}$ polku, joka alkaa apumuuttujan V määräämästä solmusta ja missä jokaista lohkoa m_k käytetään vuorollaan valitsemaan seuraava särmä.

4. Kun kaikki lohkot on käsitelty, niin viestin m hasharvo saadaan apumuuttujaan V viimeisenä tallennetun solmun nimestä.

Tarkemmin LPS-graafit määritellään Lubotzkyn, Phillipsin ja Sarnakin artikkelissa [12] ja LPS-hashfunktio Charlesin, Lauterin ja Gorenin artikkelissa [2] sivuilla 106-110, mikä sisältää myös heidän alkuperäisen kryptoanalyysinsä esittäen funktion olevan törmäysresistantti. Kuitenkin jo seuraavana vuonna ennen kuin artikkelia oli edes julkaistu painettuna Tillich ja Zémor mursivat törmäysresistanssin omassa artikkelissaan [25]. Myös Petit määrittelee hyvin tarkasti LPS-hashfunktion väitöskirjassaan [17] sivuilla 102-104 ja kertoo samalla ekvivalentista määritelmästä solmujen joukolle. Hän kokoaa tehdyn kryptoanalyysin väitöskirjassaan sivuille 147-155 ja onnistuu yhdessä Lauterin ja Quisquaterin kanssa laajentamaan Zémorin ja Tillichin hyökkäystä vuodelta 2008 niin, että se murtaa myös alkukuvaresistanssin [19]. Näin ollen LPS-hashfunktio on elävä esimerkki siitä, että laajenevan graafin tietoturvaominaisuudet eivät riitä takaamaan kyseisen hashfunktion tietoturvaa, vaan myös hashfunktion tietoturvaominaisuuksien toteutumiseen on kiinnitettävä erityistä huomiota.

LPS-hashfunktion ideaa laajentaen Petit, Lauter ja Quisquater ehdottivat laajenevaa Morgensternin hashfunktiota artikkelissaan [18] luvussa 2.3. Kuitenkin heidän oma kryptoanalyysinsä vuotta myöhemmin mursi tämänkin ehdotuksen törmäys- ja alkukuvaresistanssin, mutta tarjosi myös uuden kestävämmän muunnoksen siitä. Tästä hashfunktiosta löytyykin lisätietoa Petit'n väitöskirjasta sivuilta 104-105 sekä 155-164.

Viimeinen tässä tutkielmassa esiteltävä laajeneva hashfunktio on Pizerin hashfunktio, jota Charles, Lauter ja Goren ehdottivat samassa artikkelissa [2] kuin LPS-hashfunktiota. Nimensä se saa Pizerin laajenevien graafien perheen mukaan, jonka hän alkuperäisessä artikkelissaan [20] vuonna 1990 määrittelee kvaternioalgebran (quaternion algebra) avulla. Kuitenkin Pizerin hashfunktion määritelmissä hyödynnetään vaihtoehtoista esitystä elliptisillä käyrillä. Tarvittavat perusteet niistä löytyvät tiivistetysti esimerkiksi Petit'n väitöskirjasta sivulta 266.

Esimerkki 5.3. Vrt. [17, s. 106 ja 166] ja [2, s. 97-98]. Olkoot l ja p keskenään eri suuret alkuluvut, joista l on pieni ja p suuri. Laajenevien graafien perhe G_n muodostuu eri tavoista valita nämä alkuluvut ja sen voidaan todistaa olevan Ramanujanin

perhe. Merkitään yksittäistä kyseisen perheen graafia $\Pi_{l,p}$. Graafin $\Pi_{l,p}$ solmujen joukon muodostavat kaikki supersingulaariset elliptiset käyrät äärellisessä kunnassa $\mathbb{F}_{p^2}$. Nämä solmut voidaan nimetä yksikäsitteisesti elliptisten käyrien j-invariantin avulla. Merkitään $E(j)$ tarkoittamaan elliptistä käyrää, jonka j-invariantti on j. Solmujen $E(j_1)$ ja $E(j_2)$ välillä on särmä jos ja vain jos on olemassa l-isogeniakuvaus (l-isogeny) käyrältä $E(j_1)$ käyrälle $E(j_2)$. Näin ollen graafi $\Pi_{l,p}$ on $l+1$-säännöllinen.

Näihin Pizerin laajeneviin graafeihin perustuva hashfunktio on $P\colon M \to \mathbb{B}_s$. Usein valitaan $l = 2$ sekä p on jokin 256-bittinen alkuluku, joka on 1 modulo 12, jolloin graafi $\Pi_{2,p}$ on suuntaamaton eikä sisällä moninkertaisia särmiä. Tällöin laajeneva hashfunktio P määritellään seuraavasti. Olkoon $m \in M$ binäärilukuesitys viestistä, jonka hasharvo halutaan laskea.

1. Jaetaan viesti m yhden bitin lohkoihin $m_0, \ldots, m_{j-1}$.

2. Valitaan graafi $\Pi_{2,p}$ perheestä (G_n) ottaen huomioon, että sen kaikki solmut voidaan nimetä yksikäsitteisesti s-bittisillä kokonaisluvuilla, eli $|G| \leq 2^s$. Valitaan jokin solmu graafista $\Pi_{2,p}$ ja alustetaan se apumuuttujaan V. Alustetaan myös jokin solmun V vierussolmuista apumuuttujaan V_{-1}.

3. Jokaisella k, jolla $k \in [0, j-1]$, selvitetään solmun V vierussolmut N_1, N_2 ja N_3 2-isogeniakuvausten avulla ja valitaan lohkon m_k perusteella toinen joukosta $\{N_1, N_2, N_3\} \setminus \{V_{-1}\}$. Päivitetään apumuuttujaan $V_{-1} = V$ ja apumuuttujaan V valittu solmu.

   Näin muodostetaan graafissa $\Pi_{2,p}$ polku, joka alkaa apumuuttujan V määräämästä solmusta ja missä jokaista lohkoa m_k käytetään vuorollaan valitsemaan seuraava särmä.

4. Kun kaikki lohkot on käsitelty, niin viestin m hasharvo saadaan apumuuttujaan V viimeisenä tallennetun solmun nimestä.

Pizerin laajenevien graafien perheeseen voi tutustua tarkemmin Pizerin artikkelissa [20]. Charlesin, Lauterin ja Gorenin alkuperäisessä artikkelissa [2] sivuilla 97-106 on annettu yksityiskohtaisempi määritelmä Pizerin hashfunktiolle ja kiinnitetty erityistä huomiota graafissa kulkemiseen kohdassa 3 tarvittavien isogeniakuvausten selvittämisen osalta. Lisäksi se sisältää alkuperäisen kryptoanalyysin, jossa tarkastellaan sekä graafin että hashfunktion tietoturvaominaisuuksia, ja esitetään sen olevan törmäysresistantti. Myös Petit'n väitöskirjassa on sivuilla 106-107 esitelty Pizerin hashfunktio ja sivuilla 165-170 tarkasteltu tarkemmin sen tietoturvaa sekä hasharvojen laskennan tehokkuutta. Myös Petit'n tarkastelujen jälkeen Pizerin hashfunktio säilyttää törmäysresistanssin, mutta häviää huomattavasti laskennan nopeudessa SHA- ja ZT-hashfunktioille. Hieman vähemmän yksityiskohtia sisältävä määrittely tälle hashfunktiolle on sivuilla 53-55 Maricqin opinnäytetyössä [15] ja kansantajuinen esittely Mackenzien kirjoittamassa Science-lehden artikkelissa [14], missä on myös kuva kokonaisesta laajenevasta graafista.

Lisäksi kannattaa tutustua Jon artikkelissaan [7] esittelemään laajenevaan hashfunktioon, joka perustuu Chiun laajeneviin graafeihin [3]. ZesT on puolestaan avai-

mellinen laajeneva hashfunktio, joka on saanut vaikutteita ZT-hashfunktiosta ja jonka Maricq on esitellyt opinnäytetyössään [15] sivuilla 55-56. Tarkemmin se on käsittelyssä Petit'n väitöskirjassa [17] sivuilla 185-222. Zhupa ja Polak ehdottavat viimevuotisessa artikkelissaan avaimellista laajenevaa hashfunktiota, joka perustuu Lazebnikin ja Ustimenkon laajeneviin graafeihin [11].

Kaiken kaikkiaan laajenevat hashfunktiot tarjoavat varsin tuoreen näkökulman hashfunktioihin, sillä ensimmäisestäkin ehdotuksesta on kulunut vasta hieman alle kolmekymmentä vuotta. Teoriassa laajenevat hashfunktiot lupaavat paljon, mutta vielä ei ole käytännössä tehokasta ja turvallista laajenevaa hashfunktiota saatu luotua. Aihe kiehtookin selvästi nykypäivän tutkijoita, kuten Zhupa ja Polak, sekä kryptoanalyytikkoja ympäri maailman.

Lähteet

[1] Chabaud, F., Joux, A. *Differential collisions in SHA-0*. Springer, 1998. URL: https://doi.org/10.1007/BFb0055720.

[2] Charles, D., Lauter, K. & Goren, E. *Cryptographic hash functions from expander graphs*. Journal of Cryptology, 2009. URL: https://doi-org.libproxy.tuni.fi/10.1007/s00145-007-9002-x.

[3] Chiu, P. *Cubic Ramanujan graphs*. Springer, 1992. URL: https://doi-org.libproxy.tuni.fi/10.1007/BF01285816.

[4] Diestel, R. *Graph theory - 5th electronic edition*. Springer, 2016. URL: http://diestel-graph-theory.com.

[5] Grassl, M., Ilić, I., Magliveras, S. & Steinwand, R. *Cryptanalysis of the Tillich–Zémor hash function*. Journal of Cryptology, 2011. URL: https://doi.org/10.1007/s00145-010-9063-0.

[6] Gross, J., Yellen, J. & Zhang, P. *Handbook of graph theory, 2nd edition*. Chapman and Hall/CRC, 2013. URL: https://learning.oreilly.com/library/view/handbook-of-graph/9781439880197/.

[7] Jo, H. *Hash functions based on Ramanujan graphs*. Springer, 2017. URL: https://doi.org/10.1007/978-981-10-5065-7_4.

[8] Hoory, S., Linial, N. & Widgerson, A. *Expander graphs and their applications*. American mathematical society, 2006. URL: https://doi.org/10.1090/S0273-0979-06-01126-8.

[9] Knospe, H. *A course in cryptography*. American mathematical society, 2019.

[10] Krebs, M. & Shaheen, A. *Expander families and Cayley graphs a beginner's guide*. Oxford University Press, 2011. URL: https://ebookcentral.proquest.com/lib/tampere/detail.action?docID=845893.

[11] Lazebnik, F. & Ustimenko, V. *Explicit construction of graphs with an arbitrary large girth and of large size*. Elsevier, 1995. URL: https://doi-org.libproxy.tuni.fi/10.1016/0166-218X(94)00058-L.

[12] Lubotzky, A., Phillips, R. & Sarnak, P. *Ramanujan graphs*. Springer, 1988. URL: https://doi-org.libproxy.tuni.fi/10.1007/BF02126799.

[13] Niemelä, J. *TikZ ja PGF – ohjeita ja esimerkkejä*. URL: https://www3.uef.fi/documents/250495/2301557/niemela-tikz-ohje.pdf/5285ebad-409f-431e-8502-100c932b0975, viitattu 3.12.2020.

[14] Mackenzie, D. *Hash of the future?* American Association for the Advancment of Science, 2008. URL: https://science-sciencemag-org.libproxy.tuni.fi/content/sci/319/5869/1481.full.pdf.

[15] Maricq, A. *Application of expander graphs in cryptography*. Opinnäytetyö, Whitman College, 2014. URL: https://www.whitman.edu/documents/Academics/Mathematics/2014/maricqaj.pdf.

[16] Menezes, A., Oorschot, P. & Vanstone, S. *Handbook of applied cryptography*. CRC Press, 1997.

[17] Petit, C. *On graph-based cryptographic hash functions*. Väitöskirja, Université catholique de Louvain, 2009. URL: http://hdl.handle.net/2078.1/22375.

[18] Petit, C., Lauter, K. & Quisquarter, J. *Cayley Hashes: A class of efficient graph-based hash functions*. Tutkimusraportti, Université Catholique de Louvain, 2012. URL: http://hdl.handle.net/2078.1/119980. Saatavilla myös osoitteessahttp://homepages.ulb.ac.be/~chripeti/files/Cayley.pdf, viitattu 4.12.2020.

[19] Petit, C., Lauter, K. & Quisquater J. *Full Cryptanalysis of LPS and Morgenstern hash functions*. Springer, 2008. URL: https://doi-org.libproxy.tuni.fi/10.1007/978-3-540-85855-3_18.

[20] Pizer, A. *Ramanujan graphs and Hecke operators*. American mathematical society, 1990. URL: https://doi.org/10.1090/S0273-0979-1990-15918-X.

[21] Priyadarsini, P. *A survey on some applications of graph theory in cryptography* Taylor & Francis, 2015. URL: https://doi.org/10.1080/09720529.2013.878819.

[22] Schneier, B. *Applied cryptography: protocols, algorithms and source code in C, 20th Anniversary Edition*. Wiley, 2015. URL: https://learning.oreilly.com/library/view/applied-cryptography-protocols/9781119096726/.

[23] Stevens, M., Bursztein, E., Karpman, P., Albertini, A. & Markov Y. *The First Collision for Full SHA-1*. Springer, 2017. URL: https://doi.org/10.1007/978-3-319-63688-7_19.

[24] Stinson, D. *Cryptography theory and practice*. Chapman and Hall/CRC, 2006.

[25] Tillich, J. & Zémor, G. *Collisions for the LPS expander graph hash function*. Springer, 2008. URL: https://doi-org.libproxy.tuni.fi/10.1007/978-3-540-78967-3_15.

[26] Tillich, J & Zémor, G. *Group-theoretic hash functions*. Springer, 1994. URL: https://doi-org.libproxy.tuni.fi/10.1007/3-540-57843-9_12.

[27] Tillich, J. & Zémor, G. *Hashing with SL_2*. Springer, 1994. URL: https://doi-org.libproxy.tuni.fi/10.1007/3-540-48658-5_5.

[28] West, D. *Introduction to graph theory*. Pearson, 2018.

[29] *FIPS 180-4*. NIST, 2015. URL: https://csrc.nist.gov/publications/detail/fips/180/4/final, viitattu 3.12.2020.

[30] *FIPS 202*. NIST, 2015. URL: https://csrc.nist.gov/publications/detail/fips/202/final, viitattu 3.12.2020.

[31] *WolframAlpha*. URL: https://www.wolframalpha.com, viitattu 3.12.2020.

[32] Zémor, G. *Hash functions and graphs with large girths*. Springer, 1991. URL: https://doi-org.libproxy.tuni.fi/10.1007/3-540-46416-6_44.

[33] Zhupa, E. & Polak, M. *Keyed hash function from large girth expander graphs*. Cornell University, 2019. URL: https://arxiv.org/abs/1903.06267.