HISTORIALLISIA SALAKIRJOITUSMENETELMIÄ

Satu Soini

Pro gradu -tutkielma
Lokakuu 2020

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

TURUN YLIOPISTO
Matematiikan ja tilastotieteen laitos

SOINI, SATU: Historiallisia salakirjoitusmenetelmiä
Pro gradu -tutkielma, 44 s.
Matematiikka
Lokakuu 2020

Tässä matematiikan opettajalinjan pro gradu -tutkielmassa käsitellään historiallisesti merkittäviä salakirjoitusmenetelmiä. Tutkielma on ensisijaisesti suunnattu oppimateriaaliksi, jonka lukijalta odotetaan lukion pitkän matematiikan oppimäärän hallitsemista. Lukion pitkässä oppimäärässä keskitytään suhteellisen vähän lukuteoriaan, joten tutkielman toisessa luvussa kerrataan salakirjoitusmenetelmien kannalta olennaisia lukuteorian määritelmiä ja lauseita. Samassa luvussa esitellään myös matriisit.

Kolmannessa luvussa perehdytään ensimmäiseksi kryptografian ja kryptoanalyysin käsitteisiin, minkä jälkeen esitellään yksitellen läpi kuusi erilaista salakirjoitusmenetelmää, joista jokainen on ollut aikanaan merkityksellinen. Nämä menetelmät ovat Caesarin, Vigenerén, Hillin ja Pohlig-Hellmanin salakirjoitukset, kertolaskusalakirjoitus ja affiini salakirjoitus. Jokaisen menetelmän kohdalla annetaan myös esimerkki, miten menetelmällä salataan selkokielinen viesti ja toisaalta miten menetelmällä laadittu salakirjoitus voidaan purkaa.

Viimeisessä luvussa käydään läpi erilaisia kryptoanalyysin menetelmiä, joilla on mahdollista murtaa luvussa 3 esitellyt salakirjoitukset. Luku keskittyy tämän lisäksi pohtimaan, miten eri salakirjoitusmenetelmien turvallisuutta olisi mahdollista parantaa.

Asiasanat: kryptografia, historialliset salakirjoitusmenetelmät, salakirjoituksen murtaminen.

Sisältö

1 Johdanto

Ihmiset ovat pyrkineet salaamaan viestejään lähes yhtä kauan kuin niitä on kirjoitettukin, ja erilaisia salakirjoitusmenetelmiä on hyödynnetty niin sodissa kuin henkilökohtaisten viestienkin salaamisissa. Esimerkiksi antiikin Kreikassa salattavat viestit tatuoitiin orjien päähän. Kun orjan hiukset olivat kasvaneet tarpeeksi, hänet lähetettiin matkaan. Perillä hänen hiuksensa leikattiin, ja salainen viesti saatiin luettua. Salakirjoitusmenetelmät ovat muuttuneet valtavasti antiikin Kreikan ajoista, ja nykyään erilaiset salakirjoitusmenetelmät ovat osana lähes jokaisen arkipäivää. Niitä käytetään muun muassa pankkipalveluissa, sähköisissä allekirjoituksissa ja mobiilivarmenteissa.

Salakirjoitusmenetelmiä ja niiden murtamista tutkii aivan oma tieteenalansa, kryptologia. Kryptologiaan tutustuminen kannattaa aloittaa perehtymällä ensimmäiseksi historiallisesti merkittäviin salakirjoitusmenetelmiin. Kun aloittaa opiskelun kronologisesti, on helpompi ymmärtää, miksi ja miten kryptologia on kehittynyt ja miten aiemmat menetelmät ovat vaikuttaneet uusien menetelmien syntymiseen. Historiallisesti merkittäviä salakirjoitusmenetelmiä ovat muun muassa Caesarin, Vigenéren, Hillin ja Pohlig-Hellmanin menetelmät. Nämä ja kaksi muuta menetelmää käydään läpi kolmannessa luvussa.

Salakirjoitusmenetelmiin tutustuessa ei saa unohtaa menetelmien turvallisuuden pohtimista. Viestejä salataan, jotta ulkopuoliset eivät saisi tietää, mitä niihin on kirjoitettu. Tämä ei kuitenkaan estä ketään yrittämästä murtaa niiden salausta. Yksi tunnetuimmista salauksen murroista tapahtui toisessa maailmansodassa, kun brittiläiset mursivat saksalaisten Enigma-salauslaitteen. Salausmenetelmiä käyttäessä on siis hyvä tiedostaa, mikä on menetelmän heikkous ja voisiko menetelmän turvallisuuteen itse käyttäjänä vaikuttaa. Näitä asioita käydään tarkemmin läpi tutkielman viimeisessä luvussa, jossa esitellään myös menetelmiä, joilla voi murtaa luvussa 3 esiteltyjä salakirjoitusmenetelmiä.

2 Esitietoja

Kryptografia pohjautuu suurelta osin algebraan ja lukuteoriaan, minkä vuoksi on hyvä käydä läpi muutamia tärkeitä määritelmiä ja tuloksia ennen kuin siirtyy itse salausmenetelmiin. Tarpeellisimmista määritelmistä ja lauseista on annettu myös esimerkit, jotka helpottavat asian ymmärtämistä. Luku pohjautuu lähteisiin [3], [10] ja [12].

2.1 Eukleideen algoritmi ja Bezout'n yhtälö

Määritelmä 1. Olkoon a nollasta eroava kokonaisluku. Kokonaislukua c kutsutaan luvun a *tekijäksi*, jos ja vain jos on olemassa sellainen kokonaisluku q, että $a = qc$.

Määritelmä 2. Olkoot a ja b nollasta eroavia kokonaislukuja, joiden yhteisistä tekijöistä luku c on suurin. Tällöin lukua c kutsutaan lukujen a ja b *suurimmaksi yhteiseksi tekijäksi* ja merkitään $c = \text{syt}(a, b)$.

Määritelmä 3. Lukua 1 suurempaa kokonaislukua p kutsutaan alkuluvuksi, jos sen tekijät ovat ainoastaan luku itse ja luku 1.

Lause 1. (Jakoyhtälö). Jos a ja b ovat positiivisia kokonaislukuja, niin on olemassa yksikäsitteiset kokonaisluvut q ja r siten, että

$$a = qb + r,\ 0 \leq r < b.$$

Todistus. Olkoon joukko $A = \{a - nb \mid n \in \mathbb{N}_+,\ a - nb \geq 0\} = \{a, a-b, a-2b, a-3b, ...\}$. Joukkoon A, kuuluu näin ollen ainakin luku a. Merkitään joukon A pienintä lukua $r = a - qb$. Luvun r on oltava suurempi tai yhtä suuri kuin nolla, sillä muuten se ei kuuluisi joukkoon A.

Toisaalta $r < b$, sillä jos $r \geq b$, niin $r - b \geq 0$. Sijoittamalla tähän epäyhtälöön $r = a - qb$ saadaan

$$r - b = a - qb - b = a - (q+1)b \geq 0.$$

Koska $a - (q+1)b$ on positiivinen, se kuuluu joukkoon A. Lisäksi $a - (q+1)b$ on pienempi kuin r, jolloin se olisi joukon A pienin luku. Tästä seuraa ristiriita, sillä r on joukon A pienin alkio. Siis $0 \leq r < b$.

Osoitetaan vielä, että kokonaisluvut q ja r ovat yksikäsitteiset. Tehdään vastaoletus, että olisi olemassa lisäksi kokonaisluvut s ja t siten, että $a = sb + t$, kun $r, t \in [0, b-1]$. Tällöin

$$\begin{aligned} & qb + r = sb + t \\ \Rightarrow \quad & (q-s)b = t - r \qquad \| : b \\ \Rightarrow \quad & q - s = (t-r)/b. \end{aligned}$$

Tällöin luvun b on oltava luvun $t - r$ tekijä, sillä q ja s ovat kokonaislukuja ja näin ollen myös $q - s$ on kokonaisluku. Toisaalta $r, t \in [0, b-1]$, jolloin

$$\begin{aligned} & 0 - (b-1) \leq t - r \leq b - 1 - 0 \\ \Rightarrow \quad & -(b-1) \leq t - r \leq b - 1. \end{aligned}$$

Koska b on luvun $t - r$ tekijä, niin edellinen yhtälö on voimassa ainoastaan, kun

$$\begin{aligned} & t - r = 0b \\ \Rightarrow \quad & t = r. \end{aligned}$$

Kun $t = r$, niin

$$\begin{aligned} & (q-s)b = t - r = 0 \\ \Rightarrow \quad & q = s, \end{aligned}$$

ja näin ollen q ja r ovat yksikäsitteiset. □

Lemma 1. Jos kokonaisluku $a = qb + r,\ 0 \leq r < b$, niin silloin $\text{syt}(a, b) = \text{syt}(b, r)$.

Todistus. Olkoon syt(a, b)=c. Luku c on näin ollen lukujen a ja b yhteinen tekijä ja muodosta $a = qb + r$ seuraa, että c on myös luvun r tekijä. Siis c on lukujen b ja r yhteinen tekijä. Merkitään nyt lukujen b ja r suurinta yhteistä tekijää syt(b, r)=d, jolloin $d \geq c$. Koska d on lukujen b ja r yhteinen tekijä, sen täytyy olla myös luvun a tekijä. Siis d on lukujen a ja b yhteinen tekijä ja $d \leq c$. Ollaan siis saatu, että $d \geq c$ ja $d \leq c$, mistä seuraa, että $c = d$. □

Lause 2. (Eukleideen algoritmi) Jos a ja b ovat positiivisia kokonaislukuja ja

$$\begin{aligned}
&a = q_1 b + r_1,\ 0 \leq r_1 < b,\\
&b = q_2 r_1 + r_2,\ 0 \leq r_2 < r_1,\\
&r_1 = q_3 r_2 + r_3,\ 0 \leq r_3 < r_2,\\
&\vdots\\
&r_{n-2} = q_n r_{n-1} + r_n\\
&r_{n-1} = q_{n+1} r_n,
\end{aligned}$$

niin luku r_n =syt(a, b).

Todistus. Koska positiiviset luvut $r_1,\ r_2, ...$ pienenevät koko ajan, niin ennen pitkää $r_{n+1} = 0$. Tällöin $r_{n-1} = q_{n+1} r_n$. Lemmasta 1 saadaan, että

$$\text{syt}(a, b) = \text{syt}(b, r_1) = \text{syt}(r_1, r_2) = \cdots = \text{syt}(r_{n-1}, r_n) = r_n$$

□

Esimerkki 2.1. Etsitään lukujen 114 ja 48 suurin yhteinen tekijä Eukleideen algoritmilla,

$$\begin{aligned}
114 &= 2 \cdot 48 + 18\\
48 &= 2 \cdot 18 + 12\\
18 &= 1 \cdot 12 + 6\\
12 &= 2 \cdot 6 + 0.
\end{aligned}$$

Lukujen 114 ja 48 suurin yhteinen tekijä on 6, syt(114, 48)=6. △

Lause 3. (Bezout'n yhtälö) Jos a ja b ovat positiivisia kokonaislukuja ja syt(a, b)=c, on olemassa kokonaisluvut x ja y, niin että $ax + by = c$.

Todistus. Eukleideen algoritmista nähdään, että on lukujen a ja b suurin yhteinen tekijä c voidaan kirjoittaa muodossa $r_{n-2} - q_n r_{n-1} = c$. Korvaamalla seuraavaksi vuoron perään Eukleideen algoritmista $r_{n-2} = r_{n-3} - q_{n-1} r_{n-2}$, r_{n-3} ja niin edelleen päästään lopulta muotoon $c = ax + by$. □

Esimerkki 2.2. Lukujen 114 ja 48 suurin yhteinen tekijä on 6. Etsitään luvut x ja y niin, että Bezout'n yhtälö $114x + 48y = 6$ toteutuu. Esimerkissä 2.1 etsittiin

lukujen 114 ja 48 suurin yhteinen tekijä Eukleideen algoritmilla. Bezout'n yhtälö on helpointa muodostaa seuraamalla Eukleideen algoritmia lopusta alkuun.

$$\begin{aligned}
6 &= 18 - 1 \cdot 12 && \| \text{ sij. } 12 = 48 - 2 \cdot 18 \\
&= 18 - (48 - 2 \cdot 18) \\
&= 18 - 48 + 2 \cdot 18 \\
&= 3 \cdot 18 - 48 && \| \text{ sij.} 18 = 114 - 2 \cdot 48 \\
&= 3(114 - 2 \cdot 48) - 48 \\
&= 3 \cdot 114 - 6 \cdot 48 - 48 \\
&= 3 \cdot 114 - 7 \cdot 48
\end{aligned}$$

Bezout'n yhtälöstä nähdään suoraan, että yhtälö $114x + 48y = 6$ toteutuu, kun $x = 3$ ja $y = -7$. △

2.2 Kongruenssi

Määritelmä 4. Jos ja vain jos $a - b$ on jaollinen luvulla $m \geq 1$, niin sanotaan, että luvut a ja b ovat *kongruentit modulo m* ja merkitään $a \equiv b \pmod{m}$. Jos taas $a - b$ ei ole jaollinen luvulla m, luvut a ja b ovat *epäkongruentteja modulo m* ja merkitään $a \not\equiv b \pmod{m}$.

Lause 4. $a \equiv b \pmod{m}$, jos ja vain jos on olemassa kokonaisluku q siten, että $a = b + qm$.

Todistus. Oletetaan ensin, että $a \equiv b \pmod{m}$. Tästä seuraa, että luku $a - b$ on jaollinen luvulla m ja tällöin määritelmän 1 mukaan voidaan merkitä $a - b = qm$, ja näin ollen $a = b + qm$.

Oletetaan seuraavaksi, että on olemassa kokonaisluku q siten, että $a = b + qm$. Tällöin siis $a - b = qm$ ja määritelmän 1 mukaan $a - b$ on jaollinen luvulla m. Määritelmän 4 mukaan tällöin voidaan merkitä $a \equiv b \pmod{m}$. □

Lause 5. Jokainen kokonaisluku on kongruentti (mod m) jonkin luvuista 0, 1,..., $m - 1$ kanssa.

Todistus. Olkoon a kokonaisluku. Tällöin jakoyhtälön mukaan luku a voidaan kirjoittaa yksiselitteisesti muodossa $a = qm + r$, jossa $0 \leq r \leq m - 1$. Tällöin lauseen 4 mukaan $a \equiv r \pmod{m}$. □

Edellisestä lauseesta nähdään, että kun suoritetaan laskuja modulo m, ratkaisu on jokin luvuista 0, 1, 2,..., $m - 1$. Tämän vuoksi on kehitelty joukko $\mathbb{Z}_m$, joka määritellään alla.

Määritelmä 5. Joukko $\mathbb{Z}_m = \{0, 1, 2, ..., m - 1\}$ sisältää kaikki *pienimmät epänegatiiviset jäännökset modulo m.*

Jatkossa laskutoimituksien, jotka suoritetaan modulo m, vastaukset kuuluvat lähes poikkeuksetta joukkoon $\mathbb{Z}_m$.

Lause 6. Olkoon a ja m kokonaislukuja ja $m \geq 1$. Tällöin on olemassa kokonaisluku b, joka toteuttaa yhtälön $a \cdot b \equiv 1 \pmod{m}$, jos ja vain jos syt$(a, m)$=1. Kokonaislukua b kutsutaan tällöin luvun a *käänteisluvuksi* modulo m.

Todistus. Oletetaan ensin, että syt(a, m)=1. Lauseen 2.2 mukaan tällöin on olemassa kokonaisluvut x ja y, niin että $ax + my = 1$, mistä saadaan edelleen, että $ax - 1 = my$. Luku $ax - 1$ on siis jaollinen luvulla m ja määritelmän 4 mukaan $ax \equiv 1 \pmod{m}$. Näin ollen valitaan, että $b = x$.

Oletetaan nyt vuorostaan, että luvulla a on käänteisluku modulo m. Toisin sanoen $a \cdot b \equiv 1 \pmod{m}$ ja tällöin määritelmästä 4 saadaan, että jollain kokonaisluvulla c pätee $ab - 1 = cm$. Edelleen $ab - cm = 1$. Lauseen 2.2 mukaan syt(a, m)=1.

Näin ollen luvulla a on käänteisluku modulo m, jos ja vain jos syt(a, m)=1. □

Huomautus 1. Lauseessa 6 esitettyä käänteislukua b merkitään yleisesti a^{-1}. On kuitenkin huomattava, että moduloaritmetiikassa a^{-1} on eri asia kuin $1/a$.

Kokonaisluvun a käänteisluku a^{-1} modulo m voidaan löytää helposti Bezout'n yhtälöllä (lause 3). Lauseen 6 mukaan luvulla a on käänteisluku modulo m, jos ja vain jos syt(a, m)=1. Sijoittamalla nämä arvot Bezout'n yhtälöön saadaan $ax+my = 1$. Koska $my \equiv 0$ modulo m, luvun x on oltava luvun a käänteisluku modulo m.

Esimerkki 2.3. Etsitään luvun 9 käänteisluku modulo 26 hyödyntäen Bezout'n yhtälöä. Bezout'n yhtälö on helpointa muodostaa Eukleideen algoritmin avulla

Eukleideen algoritmi:	Bézout'n yhtälö:
$26 = 2 \cdot 9 + 8$	$1 = 9 - 8$
$9 = 1 \cdot 8 + 1$	$= 9 - (26 - 2 \cdot 9)$
$8 = 8 \cdot 1 + 0$	$= 3 \cdot 9 - 26$

Bezout'n yhtälöstä nähdään suoraan, että luvun 9 käänteisluku on 3 modulo 26. △

Määritelmä 6. Luku a kuuluu joukkoon $\mathbb{Z}_m^*$, jos ja vain jos $a \in \mathbb{Z}_m$ ja jos luvulla a on käänteisluku modulo m.

Huomautus 2. Jos laskutoimitus suoritetaan alkuluvun p modulona, sen ratkaisu a kuuluu aina joukkoon $\mathbb{Z}_p^*$, sillä syt(a, p)=1

Lause 7. (Fermat'n pieni lause) Jos p on alkuluku ja a kokonaisluku, $1 \leq a \leq p-1$, niin

$$a^{p-1} \equiv 1 \pmod{p}.$$

Todistus. Olkoon p alkuluku ja a kokonaisluku, jolle pätee $1 \leq a \leq p - 1$. Koska p on alkuluku, kokonaislukujen 1, 2, 3, ..., $p - 1$ erotukset eivät ole jaollisia luvulla p ja näin ollen ne eivät ole kongruentteja keskenään modulo p.

Yksikään luvuista a, $2a$, $3a$,..., $(p-1)a$ ei näin ollen ole jaollinen luvulla p. Lisäksi luvut a, $2a$, $3a$, ..., $(p-1)a$ eivät ole kongruentteja keskenään modulo p, sillä jos näin olisi, niin olisi olemassa kokonaisluvut r ja s, $1 \leq s < r \leq p-1$, joille pätisi $ra \equiv sa \pmod{p}$. Koska syt(a, p)=1, niin lauseen 6 mukaan luvulla a on olemassa tällöin

käänteisluku a^{-1} modulo p. Kertomalla edellä ollut kongruenssiyhtälö puolittain käänteisluvulla a^{-1} saadaan

$$r \equiv s \pmod{p},$$

mikä on ristiriita, ja näin ollen luvut $a,\ 2a,\ 3a\ , ...,\ (p-1)a$ eivät ole kongruentteja keskenään modulo p.

Koska luvut 1, 2, 3, ..., $p-1$ eivät ole kongruentteja keskenään modulo p ja luvut a, $2a$, $3a$,...,$(p-1)a$ eivät ole kongruentteja keskenään modulo p, niin on oltava

$$1 \cdot 2 \cdot 3 \cdots (p-1) \equiv a \cdot 2a \cdot 3a \cdots (p-1)a \pmod{p}.$$

Tästä saadaan

$$(p-1)! \equiv (p-1)!a^{p-1} \pmod{p}.$$

Koska $(p-1)!$ ei ole jaollinen alkuluvulla p, niin $(p-1)! \not\equiv 0 \pmod{p}$. Edellä saatu kongruenssiyhtälö voidaan siis jakaa puolittain tekijällä $(p-1)!$, ja näin saadaan Fermat'n pieni lause,

$$a^{p-1} \equiv 1 \pmod{p}.$$

□

Esimerkki 2.4. Luku 5 on alkuluku. Tällöin Fermat'n pienen lauseen mukaan luvulle 3 pätee

$$3^4 (= 81) \equiv 1 \pmod{5}.$$

△

Koska laskinten tarkkuudet eivät aina välttämättä riitä, käydään seuraavaksi läpi *neliöi ja kerro -menetelmä*, joka helpottaa korkeiden potenssien kongruenssien laskemista. Menetelmässä lasketaan kantaluvun pienempien potenssien kongruensseja, joita kertomalla keskenään ja potenssien laskusääntöjä noudattamalla saadaan selville haluttu tulos. Neliöi ja kerro -menetelmän idea käy paremmin ilmi seuraavasta esimerkistä.

Esimerkki 2.5. Lasketaan $5^{157} \pmod{54}$ neliöi ja kerro -menetelmää hyödyntäen. Neliöi ja kerro -menetelmässä lasketaan ensimmäiseksi kantaluvun neliön kongruentti halutussa modulossa,

$$5^2 = 25 \pmod{54}.$$

Saatu tulos neliöidään uudelleen ja lasketaan sen kongruentti halutussa modulossa,

$$5^4 = (5^2)^2 = 25^2 = 625 \equiv 31 \pmod{54}$$

Tätä toistetaan niin kauan kunnes kantaluvun eksponentti pysyy haluttua eksponenttia pienempänä,

$$\begin{aligned}
5^8 &= (5^4)^2 = 31^2 = 961 \equiv 43 \pmod{54} \\
5^{16} &= (5^8)^2 = 43^2 = 1849 \equiv 13 \pmod{54} \\
5^{32} &= (5^{16})^2 = 13^2 = 169 \equiv 7 \pmod{54} \\
5^{64} &= (5^{32})^2 = 7^2 = 49 \pmod{54} \\
5^{128} &= (5^{64})^2 = 49^2 = 2\,401 \equiv 25 \pmod{54}.
\end{aligned}$$

Haluttu potenssilasku saadaan nyt muokattua potenssien laskusääntöjen mukaisesti edellä laskettujen tuloksien tuloksi,

$$5^{157} = 5^{128} \cdot 5^{16} \cdot 5^8 \cdot 5^4 \cdot 5 \equiv 25 \cdot 13 \cdot 43 \cdot 31 \cdot 5 \equiv 2\,166\,125 \equiv 23 \pmod{54}.$$

Siis $5^{157} \equiv 23 \pmod{54}$. △

2.3 Matriisit

Myöhemmin käytävässä Hillin salakirjoitusmenetelmässä tarvitaan matriiseja. Matriisilaskenta on matematiikan osa-alue, jolla on monia muitakin käytännön sovellutuksia salakirjoitusmenetelmien lisäksi. Tässä alaluvussa tutustutaan kevyesti matriiseihin ja käydään läpi esitiedot, joita vaaditaan Hillin salakirjoituksen hallintaan.

Määritelmä 7. *Matriisi A* on taulukko, joka koostuu *alkioista*, jotka ovat reaalilukuja. Alkioita merkitään a_{mn}, jossa m viittaa matriisin vaakariviin, jota kutsutaan lyhyesti *riviksi*, ja n matriisin pystyriviin, jota kutsutaan *sarakkeeksi*. Tyypin $m \times n$ matriisissa on m riviä ja n saraketta. Matriisit merkitään yleisesti kaarisulkeiden sisään

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}_{m \times n}, \text{ missä } a_{11},\ a_{12}, \ldots \in \mathbb{R}.$$

Esimerkki 2.6. Matriisissa A on kaksi riviä ja kaksi saraketta, jolloin matriisi A on 2×2-matriisi. Matriisin A alkiot ovat $a_{11} = 3$, $a_{12} = 5$, $a_{21} = 2$ ja $a_{22} = 7$. Tällöin voidaan merkitä

$$A = \begin{pmatrix} 3 & 5 \\ 2 & 7 \end{pmatrix}.$$

Matriisissa B on taas kolme riviä ja neljä saraketta, jolloin matriisi B on 3×4-matriisi. Matriisin B alkiot ovat $b_{11} = 1$, $b_{12} = 8$, $b_{13} = 2$, $b_{14} = 5$, $b_{21} = 0$, $b_{22} = 5$, $b_{23} = 9$, $b_{24} = 7$, $b_{31} = 6$, $b_{32} = 7$, $b_{33} = 2$ ja $b_{34} = 3$. Tällöin merkitään

$$B = \begin{pmatrix} 1 & 8 & 2 & 5 \\ 0 & 5 & 9 & 7 \\ 6 & 7 & 2 & 3 \end{pmatrix}.$$

△

Määritelmä 8. Matriisia, jossa on yhtä monta riviä ja saraketta kutsutaan *neliömatriisiksi*.

Määritelmä 9. Matriisia, jonka kaikki alkiot ovat nollia, kutsutaan *nollamatriisiksi*.

Määritelmä 10. Neliömatriisia, jonka *päälävistäjän* eli *diagonaalin* alkiot (a_{11}, a_{22},..., a_{nn}) ovat ykkösiä ja loput alkiot ovat nollia, kutsutaan *identiteettimatriisiksi*,

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Identiteettimatriisi muistuttaa reaalilukujen ykköstä. Jos matriisi kerrotaan identiteettimatriisilla, saadaan tuloksi alkuperäinen matriisi.

Määritelmä 11. Olkoon matriisi $A = \begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$, jos $c \in \mathbb{R}$, niin

$$cA = Ac = \begin{pmatrix} ca_{11} & ca_{12} & \cdots \\ ca_{21} & ca_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

Käydään seuraavaksi läpi matriisien kertolasku. Jotta matriisien tulo AB on mahdollista laskea matriisin A vaakarivillä tulee olla saman verran alkiota kuin matriisin B sarakkeessa. Esimerkiksi 2×3-matriisin ja 3×2-matriisin tulon on määritelty, kun taas 2×3-matriisin ja 2×3-matriisin tuloa ei voida laskea. Siis

$$\underset{m\times\, \boldsymbol{n}}{A} \cdot \underset{\boldsymbol{n}\times l}{B} = \underset{m\times l}{C}$$

samat

Määritelmä 12. Jos matriisi A on $m \times n$-matriisi ja matriisi B on $n \times l$-matriisi, niin tulo AB on $m \times l$-matriisi C, missä

$$c_{ml} = a_{m1}b_{1l} + a_{m2}b_{2l} + ... + a_{mn}b_{nl}.$$

Matriisien kertolaskun määritelmä voi tuntua hankalalta ja tärkeintä onkin ainoastaan ymmärtää, miten tulo lasketaan. Seuraavat esimerkit auttavat ymmärtämään paremmin, mistä on kyse.

Esimerkki 2.7. Olkoon matriisi $A = \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}$ ja $A = \begin{pmatrix} 2 & 7 \\ 5 & 0 \end{pmatrix}$. Tulo AB on määritelty, sillä A on 2×2-matriisi ja B on 2×2-matriisi, ja matriisi AB on myös kokoa 2×2.

Matriisien tulossa voi käyttää muistisääntönä ”rivi kertaa sarake” -lausahdusta. Ensimmäiseksi kerrotaan matriisin A ylimmän rivin alkioilla matriisin B ensimmäisen sarakkeen vastaavat alkiot ja summataan ne lopuksi yhteen. Siis

$$AB = \begin{pmatrix} \boxed{3} & \boxed{2} \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \boxed{2} & 7 \\ \boxed{5} & 0 \end{pmatrix} = \begin{pmatrix} 3 \cdot 2 + 2 \cdot 5 & * \\ * & * \end{pmatrix} = \begin{pmatrix} 16 & * \\ * & * \end{pmatrix}$$

Toistetaan seuraava vaihe matriisin A ensimmäiselle riville ja matriisin B toiselle sarakkeelle,

$$AB = \begin{pmatrix} \boxed{3} & \boxed{2} \\ 1 & 4 \end{pmatrix} \begin{pmatrix} 2 & \boxed{7} \\ 5 & \boxed{0} \end{pmatrix} = \begin{pmatrix} 16 & 3 \cdot 7 + 2 \cdot 0 \\ * & * \end{pmatrix} = \begin{pmatrix} 16 & 21 \\ * & * \end{pmatrix}.$$

Tämän jälkeen toistetaan edeltävät vaiheet matriisin A toiselle riville ja matriisin B ensimmäiselle sarakkeelle

$$AB = \begin{pmatrix} 3 & 2 \\ \boxed{1} & \boxed{4} \end{pmatrix} \begin{pmatrix} \boxed{2} & 7 \\ \boxed{5} & 0 \end{pmatrix} = \begin{pmatrix} 16 & 21 \\ 1 \cdot 2 + 4 \cdot 5 & * \end{pmatrix} = \begin{pmatrix} 16 & 21 \\ 22 & * \end{pmatrix}.$$

Vastaavasti tulon AB viimeiseksi alkioksi saadaan

$$AB = \begin{pmatrix} 3 & 2 \\ \boxed{1} & \boxed{4} \end{pmatrix} \begin{pmatrix} 2 & \boxed{7} \\ 5 & \boxed{0} \end{pmatrix} = \begin{pmatrix} 16 & 21 \\ 22 & 1 \cdot 7 + 4 \cdot 0 \end{pmatrix} = \begin{pmatrix} 16 & 21 \\ 22 & 7 \end{pmatrix}.$$

Siis matriisien A ja B tuloksi saadaan matriisi $AB = \begin{pmatrix} 16 & 21 \\ 22 & 7 \end{pmatrix}$. △

Esimerkki 2.8. Olkoon matriisi A=(4 3) ja matriisi $B = \begin{pmatrix} 2 & 0 \\ 3 & 5 \end{pmatrix}$. Tulo AB on määritelty, sillä A on 1×2-matriisi ja B on 2×2-matriisi. Matriisi AB on näin ollen kokoa 1×2 ja

$$AB = (4\ 3) \begin{pmatrix} 2 & 0 \\ 3 & 5 \end{pmatrix} = (4 \cdot 2 + 3 \cdot 3 \quad 4 \cdot 0 + 3 \cdot 5) = (17 \quad 15).$$

Tuloa BA ei ole määritelty, sillä matriisin B rivillä on kaksi alkiota, mutta matriisin A sarakkeella on ainoastaan yksi alkio. △

Edellä olevasta esimerkistä huomataan, että matriisien kertolaskussa eroaa reaalilukjen tulosta, sillä laskujärjestyksellä on väliä. Poikkeuksia lukuun ottamatta $AB \neq BA$.

Määritelmä 13. Olkoon A $n \times n$-neliömatriisi. Matriisia B, kutsutaan matriisin A käänteismatriisiksi, jos

$$AB = BA = I.$$

Matriisin A käänteismatriisia merkitään A^{-1}

Lause 8. Olkoon matriisi $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Matriisin A käänteismatriisi on

$$A^{-1} = (ad - bc)^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Todistus. Jotta $A^{-1} = (ad - bc)^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ on matriisin A käänteismatriisi, sen täytyy toteuttaa määritelmän 13 ehdot. Laskemalla saadaan matriisien A ja A^{-1} tuloksi

$$\begin{aligned} AA^{-1} &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} (ad - bc)^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \quad \| \text{ määritelmä 11} \\ &= (ad - bc)^{-1} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \\ &= (ad - bc)^{-1} \begin{pmatrix} ad - bc & -ab + ba \\ cd - dc & -cb + da \end{pmatrix} \quad \| \text{ määritelmä 11} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

ja matriisien A^{-1} ja A tuloksi

$$\begin{aligned} A^{-1}A &= (ad - bc)^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= (ad - bc)^{-1} \begin{pmatrix} da - bc & db - bd \\ -ca + ac & -cb + ad \end{pmatrix} \quad \| \text{ määritelmä 11} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Koska A^{-1} täyttää määritelmän 13 ehdot, se on matriisin A käänteismatriisi. □

Huomattavaa on, että edellisen lauseen matriisilla A ei ole olemassa käänteismatriisia A^{-1}, jos $ad - bc = 0$.

Esimerkki 2.9. Etsitään matriisin $A = \begin{pmatrix} 6 & 8 \\ 2 & 3 \end{pmatrix}$ käänteismatriisi A^{-1}. Lauseesta 13 saadaan

$$A^{-1} = (6 \cdot 3 - 8 \cdot 2)^{-1} \begin{pmatrix} 3 & -8 \\ -2 & 6 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 3 & -8 \\ -2 & 6 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} & -4 \\ -1 & 3 \end{pmatrix}.$$

△

3 Salakirjoitusmenetelmiä

Tässä luvussa käydään läpi salakirjoitusmenetelmiä, jotka ovat olleet historiallisesti hyvinkin merkittäviä. Tänä päivänä useampaakaan esiteltävistä menetelmistä ei

kannata enää käyttää sellaisenaan, sillä niillä saatava salaus ei ole tarpeeksi turvallinen. Kryptografian tieteenalan ja sen perusperiaatteen ymmärtämisen kannalta on silti hyvä tietää, miten kryptografia on kehittynyt ajan kuluessa ja miten viestejä voidaan salata melko yksinkertaisinkin keinoin. Ensimmäisessä alaluvussa käydään läpi, mitä kryptografia oikeastaan on, minkä jälkeen seuraavissa alaluvuissa käsitellään yksi kerrallaan tunnettuja salakirjoitusmenetelmiä. Seuraavien lukujen päälähteinä on käytetty lähteitä [3], [4], [5], [8] ja [10].

3.1 Mitä kryptografia on?

Julkisessa viestintäkanavassa kaikki halukkaat pystyvät seuraamaan ja lukemaan siellä kulkevia viestejä. *Kryptografia* on tieteenala, joka tutkii ja kehittää tapoja, miten julkisessa viestintäkanavassa pystyy kommunikoimaan toiselle niin, että keskustelun ulkopuoliset henkilöt eivät ymmärrä lähetettyjen viestien sisältöä. Toisin sanoen kryptografia pyrkii luomaan tapoja salata viestit siten, että ainoastaan vastaanottaja ymmärtää viestin sisällön. Ulkopuoliset henkilöt pystyvät yhä lukemaan nämä salatut viestit, mutta niiden sisältö on heille lähinnä hölynpölyä.

Kryptografiassa on muodostunut perinne nimetä keskustelua käyvät henkilöt Aliceksi ja Bobiksi ja keskustelun ulkopuolinen henkilö, joka pyrkii saamaan keskustelun sisällön selville, Eve Eavesdropperiksi (salakuuntelija). Kun Alice haluaa lähettää Bobille henkilökohtaisen viestin julkista kanavaa käyttäen, hänen tulee *salata* tämä viesti Bobin kanssa ennalta sovittua *salausmenetelmää* käyttäen. Kun Bob saa Alicelta salatun viestin, hänen täytyy avata salaus eli *purkaa* salattu viesti. Jotta jotakin menetelmää voidaan kutsua salausmenetelmäksi, sen täytyy sisältää määritelmässä 14 esitellyt viisi eri asiaa.

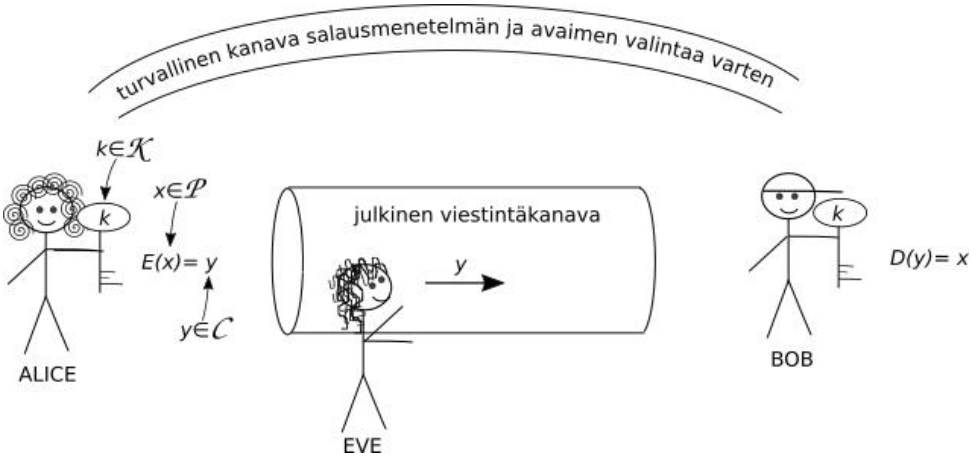
Määritelmä 14. Salausmenetelmä on viisikko $(\mathcal{P},\ \mathcal{C},\ \mathcal{K},\ E,\ D)$, missä

- $\mathcal{P}$ on joukko, jonka alkioista muodostuu *selkoteksti* (plaintext), joka halutaan salata
- $\mathcal{C}$ on joukko, johon kuuluvista alkiosta muodostuu *salakirjoitus* (ciphertext tai cryptotext), jonka viestin vastaanottaja haluaa purkaa
- $\mathcal{K}$ on joukko, joka sisältää kaikki mahdolliset *avaimet*, joilla kryptosysteemiä voidaan käyttää
- Jokaisella avaimella $k \in \mathcal{K}$ on olemassa sääntö, miten sen avulla viesti salataan. Tämä sääntö esitetään usein matemaattisesti funktiona $E : \mathcal{P} \longrightarrow \mathcal{C}$.
- Jokaisella avaimella $k \in \mathcal{K}$ on olemassa sääntö, miten sen avulla salattu viesti voidaan purkaa. Tämä sääntö esitetään usein matemaattisesti funktiona $D : \mathcal{C} \longrightarrow \mathcal{P}$. Funktio D on funktion E käänteisfunktio.

Joukko $\mathcal{P}$ voi koostua esimerkiksi yksittäisistä aakkosista tai kirjainparien yhdistelmistä, jotka saadaan muodostettua käytettävän kielen aakkosista. Salakirjoitus taas voi koostua yksittäisistä kirjaimista, kirjainpareista tai -jonoista. Toisinaan salakirjoitus voi olla myös esimerkiksi merkki- tai numerosarja. Avain taas voi olla

jokin kokonaisluku, sana tai matriisi. Kaikki seuraavissa alaluvuissa käytävät salausmenetelmät ovat *symmetrisiä salausmenetelmiä*. Symmetrisessä salausmenetelmässä käytetään samaa avainta sekä viestin salaamiseen että salakirjoituksen purkamiseen. Lisäksi ainoastaan Alice ja Bob tietävät, mikä avain on.

Even suurin toive olisi tietää, mitä Alicen ja Bobin viesteissä lukee. Häntä ei kiinnosta yksittäiset viestit, vaan hän haluaisi tietää jokaisen viestin sisällön. Jos Eve saisi tietoonsa salauksessa käytettävän avaimen, hän voisi purkaa salatut viestit vastaanottajan tavoin. Jotta Evellä on mahdollisuus päästä tavoitteeseensa hänen tulee opiskella kryptoanalyysia. *Kryptoanalyysi* on tieteenala, joka pyrkii kehittämään keinoja, joilla salakirjoitus voitaisiin *murtaa*. Toisin sanoen kryptoanalyysin keinoilla pyritään löytämään salauksessa käytetty avain. Kuvassa 1 on havainnollistettu, miten salausmenetelmä toimii ja mitkä ovat Alicen, Bobin ja Eve Eavesdropperin roolit.



Kuva 1: Alice ja Bob salaavat viestinsä, jotta Eve ei saisi selville, mistä he juttelevat. Eve tekee kuitenkin kaikkensa, jotta hän voisi salakuunnella Alicen ja Bobin käymää keskustelua.

Kryptografiassa on hyvin yleistä käyttää englannin kielen aakkostoa. Kirjaimet korvataan salausta tehdessä ja purkaessa useasti numeroilla, sillä salausmenetelmät ovat useasti hyvin matemaattisia. Yleensä kirjainta A vastaa numero 0 ja kirjainta Z numero 25. Numeroitujen aakkosten joukko on siis sama kuin joukko $\mathbb{Z}_{26}$. Taulukossa 1 on listattuna erikseen jokaista kirjainta vastaavat numerot.

Taulukko 1: Englannin kielen aakkosia vastaavat numerot.

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**
0	1	2	3	4	5	6	7	8	9	10	11	12
N	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
13	14	15	16	17	18	19	20	21	22	23	24	25

3.2 Caesarin salakirjoitus

Julius Caesar tunnetaan erityisesti muinaisen Rooman sotapäällikkönä, joka sai myöhemmin viran Rooman elinikäisenä diktaattorina (dictator perpetuus). Hänen nimestään on myös johdettu arvonimi keisari. Caesarin aikaisen Rooman juorukellona tunnettu Suetonius on kertonut, että Caesar salasi luottamukselliset viestinsä, joita hän kirjoitti poliitikoille ja ystävilleen[5]. Caesarin käyttämästä salausmenetelmästä käytetään nykyisin nimeä *Caesarin salakirjoitus* (Caesar's cipher, the shift cipher). Menetelmän nykyisestä nimestä huolimatta Caesar ei itse keksinyt käyttämäänsä salakirjoitusmenetelmää, mutta voidaan lähes kiistatta sanoa, että Caesar teki sen tunnetuksi.

Caesar muodosti salatekstin siirtämällä selkotekstin kirjaimia aakkosissa kolme kirjainta eteenpäin. Tällöin selkotekstin kirjaimesta a tulee salakirjoituksessa D ja selkotekstin kirjaimesta b vastaavasti E (selkoteksti kirjoitetaan tyypillisesti käyttäen pieniä kirjaimia ja salakirjoitus isoja kirjaimia käyttäen). Näin ollen salakirjoituksessa selkotekstin kirjain korvataan toisella kirjaimella. Tällaista salausmenetelmää kutsutaan yleisesti *korvausmenetelmäksi* (substitution cipher). Caesarin salakirjoitus voidaan esittää myös matemaattisesti, kunhan kirjaimet vaihdetaan ensin taulukon 1 mukaan niitä vastaaviksi numeroiksi. Tällöin salauksessa käytettäväksi säännöksi saadaan

$$E(x) = x + k \pmod{26}, \text{ jossa } x \in \mathcal{P} \text{ ja } k \in \mathcal{K}.$$

Joukot $\mathcal{P}$ ja $\mathcal{C}$ ovat samat joukon $\mathbb{Z}_{26}$ kanssa, sillä selkoteksti ja salakirjoitus koostuvat molemmat englannin kielen aakkosista. Mahdollisten avaimien joukko $\mathcal{K}$ ei ole pääteltävissä kuitenkaan yhtä helposti. Caesar käytti salakirjoituksissaan avaimena k lukua kolme, mutta se ei varmastikaan ole ainoa mahdollinen avain. Jos avain $k = 0$, selkoteksti ja salakirjoitus olisivat keskenään täysin samat, sillä jos mihin tahansa lukuun lisätään luku nolla luku ei muutu. Ei olisi kovinkaan järkevää käyttää avaimena lukua nolla, mutta se on silti mahdollista. Jos valitaan seuraavaksi, että avain $k = 26$, saataisiin salakirjoituksesta jälleen sama kuin selkoteksti, sillä $26 \equiv 0$ modulo 26. Esimerkiksi, kirjaimelle c saataisiin seuraavat salaukset:

$$\text{Kun } k = 0, \quad E(2) = 2 + 0 = 2 \pmod{26} \rightarrow \text{C}$$
$$\text{Kun } k = 26, \quad E(2) = 2 + 26 = 28, \; 28 \equiv 2 \pmod{26} \rightarrow \text{C}.$$

Näin ollen ei ole väliä, valitseeko avaimeksi luvun 0 vai 26. Vastaavasti avain $k = 27$ antaa saman salakirjoituksen kuin avain $k = 1$, sillä $27 \equiv 1$ modulo 26. Jatkamalla tätä päättelyketjua huomataan, että mahdolliset avaimet muodostavat joukon $\{0, 1, 2, ..., 25\}$, joka on itse asiassa sama kuin joukko $\mathbb{Z}_{26}$. Siis Caesarin salakirjoituksessa $\mathcal{K} = \mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$.

Esimerkki 3.1. Vuonna 49 eaa. yhtenä tammikuisena yönä Caesar ylitti pienen joen, Rubikonin. Joen ylityksellä oli suuri merkitys, sillä se symboloi rajaa, jonka ylittämällä Caesar julistaisi sisällissodan alkaneeksi. Ylittäessään jokea Caesar lausui ehkäpä hänen tunnetuimman lausahduksensa "Alea iacta est", joka suomennetaan usein "Arpa on heitetty". Fraasi "Alea iacta est" käännetään tyypillisesti englannin kielessä taas muotoon "The die is cast" [9]. Alice salaa tämän lausahduksen käyttäen Caesarin salakirjoitusta avaimen ollessa luku kahdeksan. Taulukon laatiminen

helpottaa viestien salaamista ja salakirjoitusten purkamista. (Lisäksi taulukkolaskentaohjelmat helpottavat ja nopeuttavat huomattavasti laskuja.)

Selkoteksti	Selkotekstin kirjainta vastaava numero x	$E(x) = x + 8 \pmod{26}$	Salakirjoitus
t	19	19+8=27, 27≡1 (mod 26)	B
h	7	7+8=15	P
e	4	4+8=12	M
d	3	3+8=11	L
i	8	8+8=16	Q
e	4	4+8=12	M
i	8	8+8=16	Q
s	18	18+8=26, 26≡0 (mod 26)	A
c	2	2+8=10	K
a	0	0+8=8	I
s	18	18+8=26, 26≡0 (mod 26)	A
t	19	1 9+8=27, 27≡1 (mod 26)	B

Selkotekstistä ”the die is cast” Alice saa Caesarin salausmenetelmällä ($k = 8$) salakirjoituksen ”BPMLQMQAKIAB”, jonka hän lähettää Bobille. △

Entä miten Bob purkaa Caesarin salausmenetelmällä salatun viestin? Caesarin salakirjoitus saatiin lisäämällä selkotekstin kirjainta vastaavaan lukuun avaimen arvo, joten näin ollen salakirjoitus puretaan vähentämällä avaimen arvo salakirjoituksen kirjainta vastaavasta luvusta.

Lause 9. Caesarin salakirjoitus voidaan purkaa funktiolla

$$D(y) = y - k \pmod{26}, \text{ jossa } y \in \mathcal{C} \text{ ja } k \in \mathcal{K}.$$

Todistus. Salakirjoitusmenetelmän määritelmän 14 mukaan salakirjoitus saadaan purettua salaukseen käytetyn funktion E käänteisfunktiolla D. Caesarin salakirjoitus muodostetaan funktiolla $E(x) = x + k \pmod{26}$. On siis varmistettava, että funktio $D(y) = y - k$ on funktion $E(x) = x + k$ käänteisfunktio. Sijoittamalla saadaan, että

$$D(E(x)) = (x + k) - k = x.$$

Näin ollen funktio D on funktion E käänteisfunktio ja Caesarin salakirjoitus saadaan purettua funktiolla D. □

Esimerkki 3.2. Bob on saanut Alicelta salakirjoituksen ”BPMLQMQAKIAB”, jonka Alice on salannut käyttäen Caesarin salakirjoitusmenetelmää. Aiemmin he olivat valinneet turvallista kanavaa käyttäen avaimeksi luvun kahdeksan. Tällöin Bob käyttää lauseen 9 mukaan salauksen purkamiseen yhtälöä $D(y) = y - 8 \pmod{26}$, missä $y \in \mathcal{C}$.

Salakirjoitus	Salakirjoituksen kirjainta vastaava numero y	$D(y) = y - 8 \pmod{26}$	Selkoteksti
B	1	1-8=-7, -7≡19 (mod 26)	t
P	15	15-8=7	h
M	12	12-8=4	e
L	11	11-8=3	d
Q	16	16-8=8	i
M	12	12-8=4	e
Q	16	16-8=8	i
A	0	0-8=-8, -8≡18 (mod 26)	s
K	10	10-8=2	c
I	8	8-8=0	a
A	0	0-8=-8, -8≡18 (mod 26)	s
B	1	1-8=-7, -7≡19 (mod 26)	t

Siis Caesarin salausmenetelmällä laadittu salakirjoitus ”BPMLQMQAKIAB” on selkokielellä, että ”the die is cast”. △

Caesarin salakirjoitus lyhyesti:

- $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{Z}_{26}$
- $E(x) = x + k \pmod{26}$, missä $x \in \mathcal{P}$ ja $k \in \mathcal{K}$
- $D(y) = y - k \pmod{26}$, missä $y \in \mathcal{C}$ ja $k \in \mathcal{K}$

3.3 Kertolaskusalakirjoitus

Caesarin salakirjoituksessa oli kyse yhteenlaskusta. Tutkitaan seuraavaksi, miten salaus toimii kertolaskun avulla. Kertolaskusalakirjoitus (multiplicative cipher, multiplication cipher), jota toisinaan kutsutaan myös kertolasku-Caesariksi, saadaan muodostettua kertomalla salattava kirjain avaimella m. Toisin sanoen

$$E(x) = mx \pmod{26}, \text{ jossa } x \in \mathcal{P} \text{ ja } m \in \mathcal{K}.$$

Joukot $\mathcal{P}$ ja $\mathcal{C}$ ovat samat joukon $\mathbb{Z}_{26}$ kanssa, sillä selkoteksti ja salakirjoitus koostuvat molemmat englannin kielen aakkosista. Ongelmaksi jää jälleen pohtia, mitkä luvut soveltuvat käytettäviksi avaimiksi.

Jos $m = 0$, niin jokaista selkotekstin kirjainta x vastaisi salakirjoituksessa kirjain A, koska mikä tahansa luku kerrottuna luvulla 0 antaa tulokseksi luvun 0. Avain $m = 0$ on siis käyttökelvoton. Jos valitaan avaimeksi $m = 1$, selkoteksti ja salakirjoitus olisivat keskenään täysin samat, koska mikä tahansa luku kerrottuna luvulla 1, on luku itse. Salausta ajatellen ei olisi mielekästä valita avaimeksi lukua 1, mutta se on toki mahdollista.

Pohditaan nyt, mitä muita avaimia ei voida käyttää avaimen $m = 0$ lisäksi. Voisiko avain m olla luku 2? Jos kerrotaan mikä tahansa luku luvulla 2, on saatu tulo parillinen ja näin ollen salakirjoituksessa ei esiinny yhtään kirjainta, jota vastaa

pariton luku. Lisäksi aina kaksi selkotekstin kirjainta salautuisivat samaksi salakirjoituksen kirjaimeksi. Näin ollen avain ei voi olla luku 2, koska silloin salakirjoitusta ei olisi mahdollista purkaa.

Esimerkki 3.3. Jos kertolaskusalakirjoituksessa valitaan avaimeksi m luku 2, kaksi selkotekstin kirjainta kuvautuisi aina samaksi salakirjoituksen kirjaimeksi. Tällöin salakirjoituksessa esiintyy tasan puolet englannin kielen aakkosista, eikä salakirjoituksen purkaminen olisi mahdollista.

Selkoteksti	Selkotekstin kirjainta vastaava numero x	$E(x) = 2x \pmod{26}$	Salakirjoitus
a	0	2·0=0	A
b	1	2·1=2	C
⋮	⋮	⋮	⋮
m	12	2·12=24	Y
n	13	2·13=26, 26≡0 (mod 26)	A
o	14	2·14=28, 28≡2 (mod 26)	C
⋮	⋮	⋮	⋮
z	25	2·25=50, 50≡ 24 (mod 26)	Y

△

Samasta syystä kuin, että avaimeksi ei voi valita lukua 2, avaimeksi ei voi valita mitään muutakaan parillista lukua. Avaimeksi ei kannata turhaan valita lukua 25 suurempaa lukua, sillä ainoastaan isommilla luvuilla laskuista tulee hankalampia. Esimerkiksi ei ole mitään merkitystä, salataanko selkoteksti käyttäen avainta 3 vai 29, sillä $29 \equiv 3$ modulo 26.

$$\text{Kun } m = 3, \quad E(2) = 3 \cdot 2 = 6 \pmod{26} \to \text{G}$$
$$\text{Kun } m = 29, \; E(2) = 29 \cdot 2 = 58, \; 58 \equiv 6 \pmod{26} \to \text{G}.$$

Nyt ollaan päästy tilanteeseen, jossa mahdollisia avaimia ovat luvut 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 ja 25. Kuitenkin näistä vielä yksi on kelvoton, ja se on luku 13. Jos $m = 13$, saadaan esimerkiksi selkotekstin kirjaimille b ja d täysin samat salakirjoitukset, mistä johtuen se ei sovellu avaimeksi:

$$\text{b}: \quad E(1) = 13 \cdot 1 = 13 \to \text{N}$$
$$\text{d}: \quad E(3) = 13 \cdot 3 = 39, \; 39 \equiv 13 \pmod{26} \to \text{N}.$$

Näin ollen kertolaskusalakirjoituksessa $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$ ja $\mathcal{K} = \{1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, 25\}$. Huomion arvoista on, että joukon $\mathcal{K}$ alkioille pätee ehto: syt$(m, 26)$=1, kun $m \in \mathcal{K}$. Toisin sanoen joukko $\mathcal{K}$ koostuu lauseen 6 mukaan niistä luvuista, joilla on käänteisluku mod 26.

Esimerkki 3.4. Alice haluaa lähettää Bobille viestin ”multiply” käyttäen kertolaskuun perustuvaa salakirjoitusta. He ovat sopineet aiemmin turvallista kanavaa käyttäen, että avain $m = 5$.

Selkoteksti	Selkotekstin kirjainta vastaava numero x	$E(x) = 5x \pmod{26}$	Salakirjoitus
m	12	5·12=60, 60≡8 (mod 26)	I
u	20	5·20=100, 100≡22 (mod 26)	W
l	11	5·11=55, 55≡3 (mod 26)	D
t	19	5·19=95, 95≡17 (mod 26)	R
i	8	5·8=40, 40≡14 (mod 26)	O
p	15	5·15=75, 75≡23 (mod 26)	X
l	11	5·11=55, 55≡3 (mod 26)	D
y	24	5·24=120, 120≡16 (mod 26)	Q

Alicen lähettämä salaviesti Bobille on ”IWDROXDQ”. △

Lause 10. Kertolaskusalakirjoituksen purkaminen tapahtuu funktiolla

$$D(y) = m^{-1}y \pmod{26}, \text{ jossa } y \in \mathcal{C} \text{ ja } m \in \mathcal{K}.$$

Todistus. Määritelmässä 14 todettiin, että salauksen purkufunktio D on salausfunktion E käänteisfunktio. Funktion $E(x) = mx \pmod{26}$ käänteisfunktio saadaan ratkaisemalla x yhtälöstä $y = mx$. Siis

$$\begin{aligned} & y = mx \quad \| \cdot m^{-1} \\ \Rightarrow \quad & x = m^{-1}y, \end{aligned}$$

Näin ollen kertolaskusalakirjoitus voidaan purkaa funktiolla $D(y) = m^{-1}y \pmod{26}$. □

On huomattava, että funktiossa D esiintyvä kerroin m^{-1} modulo 26 ei ole sama asia kuin $\frac{1}{m}$. Kerroin m^{-1} voidaan etsiä esimerkiksi Eukleideen algoritmin ja Bézout'n yhtälön avulla.

Esimerkki 3.5. Alice on lähettänyt Bobille kertolaskusalakirjoituksen ”I W D R O X D Q”. Bob tietää, että Alice on käyttänyt avainta $m = 5$. Bob aloittaa viestin purkamisen etsimällä luvun 5 käänteisluvun modulo 26 Eukleideen algoritmin ja Bezout'n yhtälön avulla:

Eukleideen algoritmi:	Bézout'n yhtälö:
$26 = 5 \cdot 5 + 1$	$1 = 26 - 5 \cdot 5$
$5 = 5 \cdot 1 + 0$	

Luvun 5 käänteisluku modulo 26 on $-5 \equiv 21 \pmod{26}$, ja näin ollen Bob saa purettua Alicen lähettämä kertolaskusalauksen funktiolla $D(y) = 21y \pmod{26}$, missä $y \in \mathcal{C}$. Bob purkaa viestin laskemalla seuraavat laskut.

Salakirjoitus	Salakirjoituksen kirjainta vastaava numero y	$D(y) = 21y \pmod{26}$	Selkoteksti
I	8	21·8=168, 168≡12 (mod 26)	m
W	22	21·22=462, 462≡20 (mod 26)	u
D	3	21·3=63, 63≡11 (mod 26)	l
R	17	21·17=357, 357≡19 (mod 26)	t
O	14	21·14=294, 294≡8 (mod 26)	i
X	23	21·23=483, 483≡15 (mod 26)	p
D	3	21·3=63, 63≡11 (mod 26)	l
Q	16	21·16=336, 336≡24 (mod 26)	y

Alicen Bobille lähettämässä viestissä luki ”multiply”. △

Kertolaskusalakirjoitus lyhyesti:

- $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$ ja $\mathcal{K} = \{m \in \mathbb{Z}_{26} \mid \text{syt}(m,\, 26){=}1\}$
- $E(x) = mx \pmod{26},$ missä $x \in \mathcal{P}$ ja $m \in \mathcal{K}$
- $D(y) = m^{-1}y \pmod{26},$ missä $y \in \mathcal{C}$ ja $m \in \mathcal{K}$

3.4 Affiini salakirjoitus

Caesarin salakirjoitus ja kertolaskusalakirjoitus ovat molemmat pohjimmiltaan melko yksinkertaisia. Alice ja Bob pelkäävät, että Eve keksii liiankin helposti, miten he salaavat viestinsä, joten he haluavat kehitellä monimutkaisemman salausjärjestelmän. Mitä tapahtuu, jos Alice ja Bob yhdistäisivätkin nämä kaksi eri salausmenetelmää yhdeksi?

Kahden salakirjoitusmenetelmän yhdistäminen ei ole kovinkaan uusi keksintö. Kuitenkin Caesarin ja kertolaskusalakirjoituksen yhdistäminen keskenään on verrattain tuore, sillä idea tästä on luultavimmin muodostunut vasta 1930-luvulla. Näiden kahden salausmenetelmän yhdistelmää kutsutaan affiiniksi salakirjoitukseksi (affine cipher). Affiini salakirjoitus on yhdistelmä kahdesta salausmenetelmästä, jolloin siinä käytettävä avain koostuu näin ollen kahdesta osasta. Yleisesti avain ilmoitetaan lukuparina kaarisulkeiden sisällä, (m, k). Ensin selkotekstin kirjain kerrotaan avaimella m, kuten kertolaskuun perustuvassa salakirjoituksessa, ja sen jälkeen saatuun lukuun lisätään avain k, kuten Caesarin salakirjoitusmenetelmässä. Luvulle m pätevät samat ehdot kuin kertolaskusalakirjoituksen avaimelle ja luvulle k taas pätevät samat ehdot kuin Caesarin salakirjoituksessa käytettävälle avaimelle. Koska molemmissa, sekä Caesar salauksessa että kertolaskuun perustuvassa salauksessa, $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$, niin ne ovat samat myös affiinissa salauksessa. Matemaattisesti ilmaistuna affiinilla salakirjoituksella salaaminen tapahtuu käyttämällä funktiota:

$$E(x) = mx + k \pmod{26},\ \text{missä } x \in \mathcal{P},\ k \in \mathbb{Z}_{26} \text{ ja } \text{syt}(m,\ 26) = 1.$$

Esimerkki 3.6. Atbash on hyvin vanha salakirjoitusmenetelmä ja sitä on muun muassa käytetty Raamatussa Jeremian kirjassa. Atbash-menetelmässä aakkosten ensimmäinen kirjain kuvautuu viimeiseksi, eli kirjaimesta A tulee kirjain Z, toinen

kirjain kuvautuu toiseksi viimeiseksi, eli kirjaimesta B tulee kirjain Y, ja niin edelleen.

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
Z	Y	X	W	V	U	T	S	R	Q	P	O	N	M	L	K	J	I	H	G	F	E	D	C	B	A

Myöhemmin on huomattu, että Atbash-salakirjoitusmenetelmä on oikeastaan affiini salausmenetelmä. Eve selvittää, mikä on tällöin affiinissa menetelmässä käytettävä avainpari (m, k).

Selkoteksti	Selkotekstin kirjainta vastaava numero x	$E(x) = mx + k \pmod{26}$	Salakirjoitus
a	0	$m \cdot 0 + k = 25 \pmod{26}$	Z
b	1	$m \cdot 1 + k = 24 \pmod{26}$	Y
c	2	$m \cdot 2 + k = 23 \pmod{26}$	X
d	3	$m \cdot 3 + k = 22 \pmod{26}$	W
$\vdots$	$\vdots$	$\vdots$	$\vdots$
w	22	$m \cdot 22 + k = 3 \pmod{26}$	D
x	23	$m \cdot 23 + k = 2 \pmod{26}$	C
y	24	$m \cdot 24 + k = 1 \pmod{26}$	B
z	25	$m \cdot 25 + k = 0 \pmod{26}$	A

Eve tekee yllä olevista taulukoiduista arvoista havainnon, että

$$\begin{aligned} E(x) + x &= 25 \qquad \| - x \\ E(x) &= -x + 25. \end{aligned}$$

Koska laskut tehdään modulo 26, niin $-1 \equiv 25 \pmod{26}$, ja tällöin funktioksi $E(x)$ saadaan

$$E(x) = 25x + 25 \pmod{26}.$$

Atbash salakirjoitusmenetelmällä saadaan sama salakirjoitus kuin affiinilla salakirjoituksella avaimen ollessa (25, 25). △

Lause 11. Affiinin salakirjoituksen purkaminen tapahtuu käyttämällä funktiota

$$D(y) = m^{-1}(y - k) \pmod{26}, \text{ missä } y \in \mathcal{C},\ k \in \mathbb{Z}_{26} \text{ ja syt}(m,\ 26) = 1.$$

Todistus. Affiinin salakirjoituksen purkamiseen käytettävä funktio D on helpointa löytää etsimällä funktion $E(x) = mx + k$ käänteisfunktio. Funktion E käänteisfunktio saadaan ratkaisemalla yhtälöstä $y = mx + k$ muuttuja x:

$$\begin{aligned} & y = mx + k \qquad \| - k \\ \Rightarrow \quad & mx = y - k \qquad \| \cdot m^{-1} \\ \Rightarrow \quad & x = m^{-1}(y - k). \end{aligned}$$

Näin ollen afiini salakirjoitus puretaan funktiolla $D(y) = m^{-1}(y - k) \pmod{26}$. □

Esimerkki 3.7. Alice on lähettänyt Bobille affiinilla salakirjoitusmenetelmällä salatun viestin ”MVXLTBJ”. Avain on (3,7). Bob purkaa Alicen lähettämän viestin käyttämällä yllä johdettua affiinin salauksen purkufunktiota $D(y)$. Ensimmäisenä Bob etsii luvun 3 käänteisluvun modulo 26:

Eukleideen algoritmi:	Bézout'n yhtälö:
$26 = 8 \cdot 3 + 2$	$1 = 3 - 2$
$3 = 1 \cdot 2 + 1$	$= 3 - (26 - 8 \cdot 3)$
$2 = 2 \cdot 1 + 0$	$= 9 \cdot 3 - 26.$

Luvun 3 käänteisluku modulo 26 on 9. Bob purkaa Alicen viestin funktiolla $D(y) = 9(y - 7) \pmod{26}$.

Salakirjoitus	Salakirjoituksen kirjainta vastaava numero y	$D(y) = 9(y-7) \pmod{26}$	Selkoteksti
M	12	9(12-7)=45, 45≡19 (mod 26)	t
V	21	9(21-7)=126, 126≡22 (mod 26)	w
X	23	9(23-7)=144, 144≡14 (mod 26)	o
L	11	9(11-7)=36, 36≡10 (mod 26)	k
T	19	9(19-7)=108, 108≡4 (mod 26)	e
B	1	9(1-7)=-54, -54≡24 (mod 26)	y
J	9	9(9-7)=18	s

Alice lähetti Bobille viestin ”two keys”. △

Affiini salakirjoitus lyhyesti:

- $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$ ja $\mathcal{K} = \{(k,\ m) \in \mathbb{Z}_{26}^2 \mid \text{syt}(m,\ 26)=1\}$
- $E(x) = mx + k \pmod{26},$ missä $x \in \mathcal{P}$ ja $(k,\ m) \in \mathcal{K}$
- $D(y) = m^{-1}(y - k) \pmod{26},$ missä $y \in \mathcal{C}$ ja $(k,\ m) \in \mathcal{K}$

3.5 Vigenéren salakirjoitus

Sähköiset salauslaitteet olivat melko suuressa roolissa toisessa maailmansodassa. Esimerkiksi saksalaisten käyttämän laitteen, Enigman, salauksen murtamisella on pohdittu olleen vaikutusta jopa sota-ajan kestoon. Toisessa maailmansodassa käytössä olleet laitteet pohjautuivat monimutkaisiin moniaakkosellisiin salauksiin (polyalphabetic cipher).

1400-1500-luvuilla kehittyi uusi salakirjoituksen muoto, joka tunnetaan nykyään moniaakkosellisena korvausmenetelmänä. Edellä käydyissä salausmenetelmissä kaikki selkotekstin kirjaimet salattiin samaa avainta käyttäen. *Moniaakkosellisissa korvausmenetelmissä* selkoteksti taas salataan paloissa käyttäen jokaiselle osioille eri avainta. Kun selkotekstin kirjaimia salataan useammalla eri avaimella, selkotekstin

kirjain ei salaudu välttämättä aina samaksi salakirjoituksen kirjaimeksi. Selkotekstin kirjain a voi salautua viestin toisessa kohtaa kirjaimeksi B ja toisessa kohtaa taas kirjaimeksi C. Tähän asti käydyt salakirjoitukset ovat olleet *monoaakkosellisia*, eli selkotekstin kirjain on salautunut aina samaksi salakirjoituksen kirjaimeksi.

Ehkäpä tunnetuin yksinkertainen moniaakkosellinen korvausmenetelmä on Vigenéren salakirjoitus. Nimi Vigénere tulee ranskalaiselta Blaise de Vigenéreltä. Vigenére innostui kryptologiasta ollessaan 26-vuotiaana Roomassa suorittamassa diplomaattista tehtävää ja jäätyään eläkkeelle hän kirjoitti kirjan "Traicte des Chiffres" (1586). Kirjassaan hän muun muassa esitteli nykyisin Vigenéren salakirjoituksena tunnetun moniaakkosellisen korvausmenetelmän. Salakirjoitus oli alkujaan italialaisen Giovan Batista Belason työn tulos, eikä Vigenére milloinkaan väittänyt keksineensä menetelmää itse. 1800-luvulla Bellasolle kuuluva kunnia laitettiin kuitenkin tuntemattomasta syystä Vigenéren nimiin, ja yhä tänä päivänä on olemassa Vigenéren salakirjoitus Bellason salakirjoituksen sijaan [5].

Vigenéren salakirjoituksella salaaminen muistuttaa hieman Caesarin salakirjoitusmenetelmää, mutta kirjaimen sijaan Vigenéren salakirjoituksessa käytetään avaimena sanaa. Jos avainsanassa on n kappaletta kirjaimia, sitä voidaan merkitä $k = (k_1, k_2, ..., k_n) \in \mathcal{K}$. Salattava selkoteksti pilkotaan avaimen pituisiin pätkiin, jolloin yhtä selkotekstin pätkää voidaan merkitä avainta vastaavalla tavalla, $x = (x_1,\ x_2, ..., x_n) \in \mathcal{P}$. Esimerkiksi jos avainsanassa on kolme kirjainta sana "vigenere" pilkotaan seuraavasti

selkoteksti:	v	i	g		e	n	e		r	e
i :	1	2	3		1	2	3		1	2

Avaimen ensimmäinen termi k_1 summataan selkotekstin pätkän ensimmäiseen termiin x_1. Avaimen toinen termi k_2 summataan selkotekstin pätkän toiseen termiin x_2 ja niin edelleen. Voidaan siis ajatella, että samassa kohtaa selkotekstin pätkää olevat kirjaimet x_i salataan Caesarin salakirjoituksella samaa avainta k_i käyttäen. Esimerkiksi pätkien ensimmäiset kirjaimet ($i = 1$) salataan avaimella k_1, jolloin salausfunktio on $E(x_1) = x_1 + k_1 \pmod{26}$. Koska salaus muodostetaan pätkissä, niin $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{Z}_{26}^n$ ja salausfunktio E voidaan kokonaisuudessaan esittää muodossa

$$E(x) = E(x_1,\ x_2, ...,\ x_n) = (x_1 + k_1,\ x_2 + k_2, ...,\ x_n + k_n) \pmod{26},$$
missä $x \in \mathcal{P},\ k \in \mathcal{K}$.

Esimerkki 3.8. Giovan Bastisa Belaso kutsui kehittelemänsä moniaakkosellisen salausmenetelmän avainta "tunnussanaksi", joka on englanniksi "countersign" [5]. Alice salaa tämän sanan käyttäen Vigenéren salakirjoitusta avainsanan ollessa "key", joka voidaan kirjoittaa myös muodossa $k = (k_1,\ k_2,\ k_3)$=(10, 4, 24). Koska avainsanassa "key" on kolme kirjainta, niin salattava sana jaetaan kolmen kirjaimen mittaisiin pätkiin:

selkoteksti:	c	o	u		n	t	e		r	s	i		g	n
i :	1	2	3		1	2	3		1	2	3		1	2

Vigenéren salauksessa ei haittaa, vaikka salattavan sanan pituus ei ole jaollinen avaimen pituudella.

Alice salaa yksi kirjain kerrallaan viestin alla olevan taulukon mukaisesti.

Selkoteksti	Selkotekstin kirjainta vastaava numero x_i	$E(x_i) = x_i + k_i \pmod{26}$	Salakirjoitus
c	2	2+10=12	M
o	14	14+4=18	S
u	20	20+24=44, 44≡18 (mod 26)	S
n	13	13+10=23	X
t	19	19+4=23	X
e	4	4+24=28, 28≡2 (mod 26)	C
r	17	17+10=27, 27≡1 (mod 26)	B
s	18	18+4=22	W
i	8	8+24=32, 32≡6 (mod 26)	G
g	6	6+10=16	Q
n	13	13+4=17	R

Alice sai selkotekstin ”countersign” salakirjoitukseksi Vigenéren salakirjoitusmenetelmällä ”MSS XXC BWG QR”. △

Koska viestin salaaminen Vigenéren salakirjoitusmenetelmällä muistutti paljon Caesarin salakirjoitusta, niin myös salattujen viestien purkamisissa on samoja vaiheita. Salakirjoitus on saatu lisäämällä selkotekstin kirjainta vastaavaan lukuun tietty avaimen arvo k_i. Jotta salakirjoitus saadaan purettua selkotekstiksi, täytyy salakirjoituksen kirjainta vastaavasta luvusta vähentää sitä vastaava avaimen termi k_i. Siis

$$D(y) = D(y_1,\ y_2, ...,\ y_n) = (y_1 - k_1,\ y_2 - k_2, ...,\ y_n - k_n) \pmod{26},$$
missä $y \in \mathcal{C},\ k \in \mathcal{K}$.

Esimerkki 3.9. Belaso esitteli nykyään Vigenéren salakirjoituksena tunnetun salausmenetelmän vuonna 1553. Tätä aiemmin ensimmäisessä painetussa kryptologian kirjassa "Polygraphiae libri sex"(1518) oli kuitenkin jo pohdittu, olisiko mahdollista toteuttaa moniaakkosellinen salausmenetelmä. Alice lähettää Bobille salakirjoitetun viestin ”PFZWDSDLQG”, jossa on salattuna ensimmäisen kryptologiaa käsittelevän kirjan kirjoittajan nimi.

Alice on salannut viestin Vigenéren salakirjoitusmenetelmällä avaimen ollessa sana "word", joka voidaan kirjoittaa muodossa $(k_1,\ k_2,\ k_3,\ k_4)$=(22, 14, 17, 3). Bob jakaa ensin salakirjoituksen neljän kirjaimen osiin:

salakirjoitus:	P	F	Z	W	D	S	D	L	Q	G
i :	1	2	3	4	1	2	3	4	1	2

Edellisen jaotteluun pohjautuen Bob purkaa viestin salauksen. Purkamisen avuksi hän muodostaa alla olevan taulukon.

Salakirjoitus	Salakirjoituksen kirjainta vastaava numero y_i	$D(y_i) = y_i - k_1 \pmod{26}$	Selkoteksti
P	15	15-22=-7, -7≡19 (mod 26)	t
F	5	5-14=-9, -9≡17 (mod 26)	r
Z	25	25-17=8	i
W	22	22-3=19	t
D	3	3-22=-19, -19≡7 (mod 26)	h
S	18	18-14=4	e
D	3	3-17=-14, -14≡12 (mod 26)	m
L	11	11-3=8	i
Q	16	16-22=-6, -6≡20 (mod 26)	u
G	6	6-14=-8, -8≡18 (mod 26)	s

Ensimmäisen painetun kryptologiaa käsittelevän kirjan kirjoitti saksalainen Trithemius [5]. △

Vigenéren salakirjoitus lyhyesti:

- $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{Z}_{26}^n$
- $E(x_1,\ x_2, ...,\ x_n) = (x_1 + k_1,\ x_2 + k_2, ...,\ x_n + k_n) \pmod{26}$, missä $x_i \in \mathcal{P},\ k_i \in \mathcal{K}$
- $D(y_1,\ y_2, ...,\ y_n) = (y_1 - k_1,\ y_2 - k_2, ...,\ y_n - k_n) \pmod{26}$, missä $y_i \in \mathcal{C},\ k_i \in \mathcal{K}$

3.6 Hillin salakirjoitus

Lester S. Hill oli yhdysvaltalainen matemaatikko, joka tunnetaan erityisesti työstään kryptografian parissa. Vuoden 1929 kesällä Hill julkaisi seitsemänsivuisen artikkelin "Cryptography in an Algebraic Alphabet"[2]. Kyseisessä artikkelissa Hill esitteli moniaakkosellisen salakirjoitusmenetelmän, jota nykyisin kutsutaan Hillin salakirjoitukseksi (Hill's cipher). 1500-luvulla oli esitelty idea menetelmästä, jonka avulla pystyi salaamaan yhdellä kertaa kaksi kirjainta, ja vuosisatoja myöhemmin 1800-luvulla kehiteltiin siihen toimiva menetelmä. Hillin kehittelemässä menetelmässä oli uutta aiempiin salakirjoitusmenetelmiin verrattuna, että sen avulla pystyy salaamaan yhdellä kertaa niin monta kirjainta samanaikaisesti kuin vain haluaa. Hillin edistyksellinen idea oli muotoilla avain matriisimuotoon, mikä mahdollisti menetelmän hienouden, useamman kirjaimen samanaikaisesti salaamisen.

Käydään seuraavaksi tarkemmin läpi Hillin salakirjoitus, jossa salataan samanaikaisesti kaksi kirjainta yhdellä kertaa. Salattava sana pilkotaan kahden kirjaimen pätkiin, joita merkitään $(x_1,\ x_2) \in \mathcal{P}$ ja tällöin $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}^2$. Hillin salakirjoituksessa on tärkeää, että jokainen selkotekstin pätkä on määrätyn mittainen. Kun salataan kahta kirjainta samanaikaisesti, salattavassa sanassa on oltava parillinen määrä kirjaimia. Jos näin ei kuitenkaan ole voidaan sanan perään lisätä yksi vapaavalintainen kirjain, esimerkiksi x. Hillin salakirjoituksella ei siis pysty salaamaan sanaa ”key”, mutta sana ”keyx” taas voidaan.

Samoin kuin Vigenéren salakirjoitusmenetelmässä, myös Hillin menetelmässä avain on jokin sana. Kun salataan kahta kirjainta $(x_1,\ x_2)$ samaan aikaan, niin avain on neljäkirjaiminen sana. Jos taas salattaisiin kolmea kirjainta samanaikaisesti avainsanan pituus olisi $3 \cdot 3 = 9$ ja niin edelleen. Merkitään nyt avainsanan kirjaimia vastaavia numeroita termeillä $k_1,\ k_2,\ k_3$ ja k_4 ja $k_i \in \mathbb{Z}_{26}$. Salaamista varten avainsana muokataan matriisiksi $A = \begin{pmatrix} k_1 & k_2 \\ k_3 & k_4 \end{pmatrix}$. Myöhemmin nähdään, että Hillin salakirjoituksen purkuun tarvitaan matriisin A käänteismatriisia. Tästä johtuen avainsanaa valitessa on varmistettava, että matriisi A toteuttaa ehdon syt$(k_1k_4 - k_2k_3,\ 26)$=1. Hillin salakirjoitus muodostetaan matriisien kertolaskulla seuraavalla tavalla:

$$E(x_1,\ x_2) = (x_1\ x_2)A \pmod{26},\ \text{missä } (x_1,\ x_2) \in \mathcal{P},\ A = \begin{pmatrix} k_1 & k_2 \\ k_3 & k_4 \end{pmatrix} \text{ ja}$$

$$\text{syt}(k_1k_4 - k_2k_3, 26) = 1,\ \text{kun } k_i \in \mathbb{Z}_{26}.$$

Esimerkki 3.10. Alice haluaa lähettää Bobille viestin ”a matrix”. Hän ja Bob ovat päättäneet käyttää viestittelyyn Hillin salakirjoitusmenetelmää avaimella ”hill”. Alice korvaa avaimen kirjaimet numeroilla ja muodostaa niistä matriisin

$$A = \begin{pmatrix} 7 & 8 \\ 11 & 11 \end{pmatrix}.$$

Alice haluaa vielä varmistua, että valittu avain on kelvollinen, joten hän laskee

$$k_1k_4 - k_2k_3 = 7 \cdot 11 - 8 \cdot 11 = -11, \quad -11 \equiv 15 \pmod{26},$$

ja tarkistaa tämän jälkeen Eukleideen algoritmilla, että syt$(k_1k_4 - k_2k_3,\ 26)$=1,

$$\begin{aligned} 26 &= 1 \cdot 15 + 11 \\ 15 &= 1 \cdot 11 + 4 \\ 11 &= 2 \cdot 4 + 3 \\ 4 &= 1 \cdot 3 + 1 \\ 3 &= 3 \cdot 1 + 0, \end{aligned}$$

mistä nähdään suoraan, että syt(15, 26)=1. Avain ”hill” on näin ollen pätevä Hillin salakirjoitukseen. Alice jakaa salattavan sanan ”a matrix” kahden kirjaimen pituisiin pätkiin:

selkoteksti:	a	m	a	t	r	i	x	x
kirjainta vastaava numero:	0	12	0	19	17	8	23	23
i :	1	2	1	2	1	2	1	2

Koska selkotekstissä ”a matrix” on yhteensä seitsemän kirjainta, Alicen pitää lisätä vielä yksi kirjain salattavan viestin loppuun. Hän lisää kirjaimen x. Nyt Alice voi

salata viestinsä:

$$E(0,\ 12) = (0\ 12)\begin{pmatrix} 7 & 8 \\ 11 & 11 \end{pmatrix} = (0\cdot 7 + 12\cdot 11 \quad 0\cdot 8 + 12\cdot 11) = (132\ 132),$$
$$(132\ 132) \equiv (2\ 2) \pmod{26} \to \text{CC}$$

$$E(0,\ 19) = (0\ 19)\begin{pmatrix} 7 & 8 \\ 11 & 11 \end{pmatrix} = (0\cdot 7 + 19\cdot 11 \quad 0\cdot 8 + 19\cdot 11) = (209\ 209),$$
$$(209\ 209) \equiv (1\ 1) \pmod{26} \to \text{BB}$$

$$E(17,\ 8) = (17\ 8)\begin{pmatrix} 7 & 8 \\ 11 & 11 \end{pmatrix} = (17\cdot 7 + 8\cdot 11 \quad 17\cdot 8 + 8\cdot 11) = (207\ 224),$$
$$(207\ 224) \equiv (25\ 16) \pmod{26} \to \text{ZQ}$$

$$E(23,\ 23) = (23\ 23)\begin{pmatrix} 7 & 8 \\ 11 & 11 \end{pmatrix} = (23\cdot 7 + 23\cdot 11 \quad 23\cdot 8 + 23\cdot 11) = (414\ 437),$$
$$(414\ 437) \equiv (24\ 21) \pmod{26} \to \text{YV}$$

Alice saa Hillin salakirjoitusmenetelmällä ja avaimella ”hill” selkotekstin ”a matrix” salakirjoitukseksi ”CCBBZQYV”. △

Edellä olleesta esimerkistä nähdään hyvin, että Hillin salakirjoitusmenetelmä on moniaakkosellinen, sillä selkotekstin kirjain a kuvautuu ensimmäisen kerran salakirjoituksen kirjaimeksi C ja toisella kertaa salakirjoituksen kirjaimeksi B.

Lause 12. Hillin salakirjoitus puretaan funktiolla

$$D(y_1,\ y_2) = (y_1\ y_2)A^{-1} \pmod{26},\ \text{missä}\ (y_1,\ y_2) \in \mathcal{C},\ A = \begin{pmatrix} k_1 & k_2 \\ k_3 & k_4 \end{pmatrix}\ \text{ja}$$
$$\text{syt}(k_1k_4 - k_2k_3, 26) = 1,\ \text{kun},\ k_i \in \mathbb{Z}_{26}.$$

Todistus. Hillin salaukseen käytetyn funktion $E(x_1,\ x_2) = (x_1\ x_2)A$ käänteisfunktio D saadaan ratkaisemalla pari $(x_1\ x_2)$ yhtälöstä

$$\begin{aligned} & (x_1\ x_2)A = (y_1\ y_2) && \| \cdot A^{-1}\ \text{(oik.)} \\ \Rightarrow \quad & (x_1\ x_2) = (y_1\ y_2)A^{-1}. \end{aligned}$$

Näin ollen Hillin salakirjoitus puretaan funktiolla $D(y_1,\ y_2) = (y_1\ y_2)A^{-1} \pmod{26}$. □

On huomattava, että Hillin salakirjoituksen purkamiseen tarvittava käänteismatriisi A^{-1} muodostetaan lauseen 8 mukaan, mutta luku $(ad-bc)^{-1}$ on nyt käänteisluku mod 26.

Esimerkki 3.11. Bob vastaanotti Alicelta Hillin salakirjoitusmenetelmällä salatun viestin ”CCBBZQYV”. Salaukseen on käytetty avainta ”hill”, josta muodostuu

avainmatriisi $\begin{pmatrix} 7 & 8 \\ 11 & 11 \end{pmatrix}$. Matriisin A käänteismatriisi voidaan laskea kaavalla

$$A^{-1} = (k_1k_4 - k_2k_3)^{-1} \begin{pmatrix} k_4 & -k_2 \\ -k_3 & k_1 \end{pmatrix}.$$

Bob aloittaa viestin purkamisen etsimällä kertoimen $(k_1k_4 - k_2k_3)$=15 käänteisluvun.

Eukleideen algoritmi:

$$\begin{aligned} 26 &= 1 \cdot 15 + 11 \\ 15 &= 1 \cdot 11 + 4 \\ 11 &= 2 \cdot 4 + 3 \\ 4 &= 1 \cdot 3 + 1 \\ 3 &= 3 \cdot 1 + 0, \end{aligned}$$

Bézout'n yhtälö:

$$\begin{aligned} 1 &= 4 - 3 \\ &= 4 - (11 - 2 \cdot 4) \\ &= 3 \cdot 4 - 11 \\ &= 3(15 - 11) - 11 \\ &= 3 \cdot 15 - 4 \cdot 11 \\ &= 3 \cdot 15 - 4(26 - 15) \\ &= 7 \cdot 15 - 4 \cdot 26, \end{aligned}$$

mistä nähdään suoraan, että $(k_1k_4 - k_2k_3)^{-1}$ on 7. Bob saa salauksen purkamiseen tarvittavan matriisin A käänteismatriisiksi

$$A^{-1} = 7 \begin{pmatrix} 11 & -8 \\ -11 & 7 \end{pmatrix} = \begin{pmatrix} 77 & -56 \\ -77 & 49 \end{pmatrix} \equiv \begin{pmatrix} 25 & 22 \\ 1 & 23 \end{pmatrix} \pmod{26}.$$

Salakirjoitus voidaan näin ollen purkaa yhtälöllä

$$D(y_1,\ y_2) = (y_1\ y_2) \begin{pmatrix} 25 & 22 \\ 1 & 23 \end{pmatrix} \pmod{26}.$$

Ennen yhtälöön sijoittamista Bob jakaa salakirjoituksen kahden kirjaimen pätkiin ja merkitsee ylös jokaista salakirjoituksen kirjainta vastaavan numeron.

salakirjoitus:	C	C	B	B	Z	Q	Y	V
kirjainta vastaava numero:	2	2	1	1	25	16	24	21
i :	1	2	1	2	1	2	1	2

Nyt Bob voi suorittaa seuraavat salakirjoituksen purkamiseen tarvittavat laskut ja

selvittää näin, mitä Alicen viestissä lukee.

$$D(2,\ 2) = (2\ 2)\begin{pmatrix} 25 & 22 \\ 1 & 23 \end{pmatrix} = (2\cdot 25 + 2\cdot 1 \quad 2\cdot 22 + 2\cdot 23) = (52\ 90),$$
$$(52\ 90) \equiv (0\ 12) \pmod{26} \to \text{am}$$

$$D(1,\ 1) = (1\ 1)\begin{pmatrix} 25 & 22 \\ 1 & 23 \end{pmatrix} = (1\cdot 25 + 1\cdot 1 \quad 1\cdot 22 + 1\cdot 23) = (26\ 45),$$
$$(26\ 45) \equiv (0\ 19) \pmod{26} \to \text{at}$$

$$D(25,\ 16) = (25\ 16)\begin{pmatrix} 25 & 22 \\ 1 & 23 \end{pmatrix} = (25\cdot 25 + 16\cdot 1 \quad 25\cdot 22 + 16\cdot 23) = (641\ 918),$$
$$(641\ 918) \equiv (17\ 8) \pmod{26} \to \text{ri}$$

$$D(24,\ 21) = (24\ 21)\begin{pmatrix} 25 & 22 \\ 1 & 23 \end{pmatrix} = (24\cdot 25 + 21\cdot 1 \quad 24\cdot 22 + 21\cdot 23) = (621\ 1011),$$
$$(621\ 1011) \equiv (23\ 23) \pmod{26} \to \text{xx}$$

Bob sai Alicen lähettämäksi viestiksi ”a matrixx”. Koska ”a matrixx” ei ole englannin kielen sana, Bob poistaa saamastaan sanasta viimeisen x-kirjaimen pois ja saa näin viestiksi sanan ”a matrix”. △

Hillin salakirjoitus lyhyesti:

- $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}^2$ ja $\mathcal{K} = \{(k_1,\ k_2,\ k_3,\ k_4) \in \mathbb{Z}_{26}^4 \mid \text{syt}(k_1k_4 - k_2k_3,\ 26)=1\}$
- $E(x_1,\ x_2) = (x_1\ x_2)\begin{pmatrix} k_1 & k_2 \\ k_3 & k_4 \end{pmatrix} \pmod{26}$, missä $(x_1,\ x_2) \in \mathcal{P}$ ja $k_i \in \mathcal{K}$
- $D(y_1,\ y_2) = (y_1\ y_2)(k_1k_4 - k_2k_3)^{-1}\begin{pmatrix} k_4 & -k_2 \\ -k_3 & k_1 \end{pmatrix} \pmod{26}$, missä $(y_1,\ y_2) \in \mathcal{C}$ ja $k_i \in \mathcal{K}$

3.7 Pohlig-Hellmanin salakirjoitus

1970-luvulla Stephen Pohlig ja Martin Hellman alkoivat suunnitella ensimmäistä salakirjoitusmenetelmää, joka perustuisi julkiseen avaimeen. Julkinen avain salakirjoituksissa mullistaisi koko kryptografian, sillä silloin Alicen ja Bobin ei tarvitsisi käyttää enää turvallista kanavaa avaimen sopista varten, vaan he voisivat ilmoittaa toisilleen julkista kanavaa käyttäen, mitä avainta toisen tulisi milloinkin käyttää [4]. Ongelmana oli kuitenkin salakirjoituksen turvallisuuden varmistaminen. Julkisen avaimen salausmenetelmän kehittelyn lomassa vuonna 1978 Pohlig ja Hellman julkaisivat potenssilaskuun perustuvan salakirjoitusmenetelmän, joka on nimetty

kehittelijöidensä mukaan Pohlig-Hellmanin salakirjoitukseksi [6]. Tämä salausmenetelmä ei hyödynnä vielä julkista avainta, mutta menetelmä loi pohjaa julkiseen avaimeen perustuville salausmenetelmille.

Tähän mennessä käydyissä salakirjoituksissa kirjaimet on muutettu numeroiksi taulukon 1 mukaan, niin että A on 0, B on 1 ja niin edelleen. Pohlig-Hellmanin salakirjoitusmenetelmässä kirjaimia merkitään hieman toisin, jokaista kirjainta tulee vastata kaksinumeroinen luku. Tämä vaatimus saadaan täytettyä lisäämällä 0 lukujen 1-9 eteen. Tällöin siis A on 00, B on 01 ja J on 09. Kirjainten K-Z numeroinnit pysyvät yhä samoina kuin taulukossa 1.

Esimerkki 3.12. Sanat ”safe prime” merkitään Pohlig-Hellmanin salakirjoituksessa numeroina seuraavasti:

s	a	f	e	p	r	i	m	e
18	00	05	04	15	17	08	12	04

△

Pohlig-Hellmanin salakirjoitusmenetelmä on moniaakkosellinen ja sillä pystytään salaamaan useampia kirjaimia yhtäaikaisesti. Salaus muodostetaan potenssilaskulla jonkun alkuluvun p modulossa, jolloin $\mathcal{P} = \mathcal{C} = \mathbb{Z}_p$. Avain k on taas jokin luku, joka on lukujen 1 ja $p-1$ väliltä ja jolle pätee ehto syt$(k,\ p-1)$=1. Avainta k kutsutaan myös toisinaan salauseksponentiksi (encryption exponent) ja Pohlig-Hellmanin salakirjoitusmenetelmässä sitä käytetään seuraavalla tavalla:

$$E(x) = x^k \pmod{p},\ \text{missä } x \in \mathcal{P},\ \text{syt}(k,\ p-1) = 1 \text{ ja } p \text{ on alkuluku.}$$

Pohlig-Hellmanin salakirjoitusmenetelmällä voidaan siis salata useampia kirjaimia samanaikaisesti, mutta keskitytään seuraavaksi tilanteeseen, jossa salataan kerrallaan vain kahta kirjainta. Pohligin ja Hellmanin kehittelemässä menetelmässä yhdistetään kaksi salattavaa kirjainta yhdeksi. Esimerkiksi peräkkäisistä kirjaimista A ja B saadaan salattava luku (0)101 ja kirjaimista N ja O saadaan vastaavasti luku 1 314. Kun salataan kaksi kirjainta kerrallaan salattavassa tekstissä tulee olla parillinen määrä kirjaimia, jos näin ei kuitenkaan ole, lisätään tekstin perään vielä yksi kirjain, esimerkiksi kirjain x.

Esimerkki 3.13. Pohlig-Hellman salakirjoituksessa kirjaimet salataan pareittain, jolloin esimerkin 3.12 sanat ”safe prime” salattaisiin seuraavanlaisissa paloissa:

sa	fe	pr	im	ex
1800	0504	1517	0812	0423

Koska sanoissa ”safe prime” on yhteensä yhdeksän kirjainta, on sanojen perään liitetty vielä kirjain x.

△

Kun salataan kahta kirjainta kerrallaan Pohlig-Hellmanin menetelmällä, suurin salattava luku voi olla 2 525, joka vastaa kirjainparia ZZ. Tästä johtuen alkuluku p, jonka modulossa salaukseen tarvittavat laskut suoritetaan, täytyy olla lukua 2 525 suurempi. Esimerkiksi luku 2 579 on ensimmäinen lukua 2 525 suurempi alkuluku,

ja näin ollen laskut voidaan suorittaa modulo 2 579. Alkuluvuksi p olisi voinut valita jonkun muunkin ja myöhemmin huomataankin, että salauksen turvallisuuden kannalta olisi paras valita alkuluvuksi mahdollisimman suuri luku. Jotta salauksen muodostaminen ja purkaminen olisi kuitenkin nyt helpompaa, suoritetaan laskut modulossa 2 579.

Koska salauksessa suoritettavien laskujen tulokset kuuluvat alkuluvun p jäännösluokkiin, salakirjoituksen osat kuuluvat joukkoon $\mathcal{C} = \mathbb{Z}_p$. Tällöin ei kuitenkaan ole aina mahdollisuutta muuttaa salakirjoitusta kirjaimiksi. Esimerkiksi salakirjoituksessa voi esiintyä luku 2 578, sillä se kuuluu joukkoon $\mathbb{Z}_{2\ 579}$. Sen kaksi ensimmäistä numeroa, 25, vastaisivat kirjainta Z, mutta luku 78 ei vastaa mitään englannin kielen kirjainta. Tästä syystä Pohlig-Hellmanin salakirjoitus ei koostu kirjaimista, vaan se on numerosarja.

Esimerkki 3.14. Alice haluaa lähettää Bobille viestin ”safe prime”. Hän salaa sen käyttämällä Pohlig-Hellmanin salakirjoitusmenetelmää. Alice ja Bob ovat valinneet alkuluvuksi $p = 2\ 579$ ja avaimeksi salauseksponentin $k = 27$. Alice saa esimerkistä 3.12 salaukseen tarvittavat selkotekstin kirjainpareja vastaavat luvut. Koska isojen lukujen potensseja on välillä hyvin vaikea laskea Alice hyödyntää salauksessa neliöi ja kerro -menetelmää. Esimerkiksi kirjainparia ”sa” vastaa luku x=1 800, jolloin

$$\begin{aligned}
x^2 &= 1800^2 = 3\ 240\ 000 \equiv 776 \pmod{2\ 579},\\
x^4 &= (x^2)^2 = 776^2 = 602\ 176 \equiv 1\ 269 \pmod{2\ 579},\\
x^8 &= (x^4)^2 = 1\ 269^2 = 1\ 610\ 361 \equiv 1\ 065 \pmod{2\ 579},\\
x^{16} &= (x^8)^2 = 1\ 065^2 = 1\ 134\ 225 \equiv 2\ 044 \pmod{2\ 579}.
\end{aligned}$$

Käyttämällä yllä olevia tuloksia Alicen on nyt helpompi laskea, että

$$\begin{aligned}
E(x) = x^{27} &= x^{1+2+8+16} = x^1x^2x^8x^{16}\\
&\equiv \underbrace{1\ 800 \cdot 776}_{1396800\equiv 1561} \cdot \underbrace{1\ 065 \cdot 2\ 044}_{2176860\equiv 184}\\
&\equiv \underbrace{1\ 561 \cdot 184}_{287224}\\
&\equiv 955 \pmod{2\ 579}.
\end{aligned}$$

Alice salaa loput viestin kirjainpareista vastaavalla tavalla ja saa muodostettua näin alla olevan taulukon. Huomaa, että kaikki taulukon tulokset ovat annettu modulo 2 579 ja että jokaisessa salakirjoituksen pätkässä on oltava neljä numeroa.

Kirjain-pari	x	x^2	x^4	x^8	x^{16}	$E(x) = x^{27} = x^1x^2x^8x^{16}$
sa	1 800	776	1 269	1 065	2 044	0955
fe	504	1 274	885	1 788	1 563	2188
pr	1 517	821	922	1 593	2 492	1687
im	812	1 699	700	2 569	100	1628
ex	423	978	2 254	2 465	101	1018

Alice lähettää Bobille salakirjoitetun viestin ”09552188168716281028”. △

Alice on lähettänyt Bobille Pohlig-Hellmanin salakirjoitusmenetelmällä salatun viestin. Miten Bobin tulisi purkaa se? Tavallisesti potenssiyhtälöt ratkaistaan ottamalla yhtälöstä potenssia vastaava juuri. Esimerkiksi ottamalla yhtälöstä $x^4 = 81$ neljäs juuri, saadaan ratkaisuksi $x = \sqrt[4]{81} = 3$. Koska salauksessa on kuitenkin käytetty jakojäännösluokkia, niin salauksen purkaminen ei onnistu juurien avulla. Esimerkissä 3.14 Alice sai kirjaimille pr ($x = 1\ 517$) salakirjoitukseksi numerosarjan 1687. Jos otetaan tästä tuloksesta 27. juuri, saadaan $\sqrt[27]{1\ 687} \approx 1,3168$, joka ei selvästikään ole haluttu tulos 1 517. Jotta Bob saa Pohlig-Hellmanin salakirjoitusmenetelmällä salatun viestin purettua, hänen täytyy hyödyntää Fermat'n pientä lausetta.

Lause 13. Pohlig-Hellmanin salakirjoitus puretaan funktiolla

$$D(y) = y^{k^{-1}} \pmod{p}, \text{ missä } y \in \mathcal{C} \text{ ja } k \cdot k^{-1} \equiv 1 \pmod{p-1}.$$

Todistus. Fermat'n pienessä lauseessa sanotaan, että $x^{p-1} \equiv 1 \pmod{p}$, kun p on alkuluku, ja se voidaan kirjoittaa myös muodossa

$$x^{p-1} \equiv x^0 \pmod{p}.$$

Vaikka siis Fermat'n pieni lause lasketaan modulo p, niin eksponenttiin liittyvät laskut suoritetaan modulo $p-1$, sillä $p - 1 \equiv 0 \pmod{p-1}$.

Pohlig-Hellmanin salakirjoitus y saadaan korottamalla salattava luku x potenssiin k. Siis $x^k = y \pmod{\text{p}}$. Salauseksponentille k määrättiin aiemmin ehto, että syt$(k,\ p-1)$=1. Tästä ehdosta seuraa, että luvulla k on käänteisluku k^{-1} modulo $p-1$. Toisin sanoen

$$k \cdot k^{-1} \equiv 1 \pmod{p-1}.$$

Fermat'n pienestä lauseesta taas saatiin, että eksponenttiin liittyvät laskutoimitukset suoritetaan modulossa $p-1$. Näin ollen yhtälö $x^k = y \pmod{\text{p}}$, voidaan korottaa puolittain salauseksponentin k käänteisluvulla k^{-1} modulo $p-1$,

$$\begin{aligned} & x^k = y && \|()^{k^{-1}} \\ \Rightarrow \quad & (x^k)^{k^{-1}} = y^{k^{-1}} \\ \Rightarrow \quad & x^{k \cdot k^{-1}} = y^{k^{-1}} \\ \Rightarrow \quad & x = y^{k^{-1}}. \end{aligned}$$

Tästä nähdään, että salaukseen käytetyn funktion $E(x) = x^k$ käänteisfunktio on

$$D(y) = y^{k^{-1}} \pmod{p}, \text{ missä } y \in \mathcal{C} \text{ ja } k \cdot k^{-1} \equiv 1 \pmod{p-1}.$$

□

Pohlig-Hellmanin salauksen purkuun tarvittavaa avaimen käänteislukua kutsutaan toisinaan *purkueksponentiksi* (decryption exponent) ja se saadaan etsittyä esimerkiksi Eukleideen algoritmia ja Bézout'n yhtälöä hyödyntämällä.

Esimerkki 3.15. Bob sai Alicelta Pohlig-Hellmanin menetelmällä salatun viestin ”09550537168716281028”. Bob tietää, että salauksessa on käytetty salauseksponenttia 27 ja alkuluvuksi p on valittu 2 579. Ensimmäiseksi Bobin tulee etsiä luvun 27 käänteisluku modulo 2 578

Eukleideen algoritmi:	Bézout'n yhtälö:
$2\,578 = 95 \cdot 27 + 13$	$1 = 27 - 2 \cdot 13$
$27 = 2 \cdot 13 + 1$	$= 27 - 2(2\,578 - 95 \cdot 27)$
$13 = 13 \cdot 1 + 0$	$= 191 \cdot 27 - 2 \cdot 2\,578.$

Bob saa luvun 27 käänteisluvuksi 191 modulo 2 578. Näin ollen hän purkaa Alicen muodostaman salauksen yhtälöllä

$$D(y) = y^{191} \pmod{2579}, \text{ missä } y \in \mathcal{C}.$$

Ennen yhtälöön sijoittamista Bob jakaa salakirjoituksen neljän numeron pätkiin

$$0955 \quad 2188 \quad 1687 \quad 1628 \quad 1018,$$

jotka hän purkaa yksi kerrallaan. Esimerkiksi ensimmäiseksi Bob tutkii lukua $y = 955$, jolle

$$\begin{aligned}
&y^2 = 955^2 = 912\,025 \equiv 1\,638 \pmod{2\,579},\\
&y^4 = (y^2)^2 = 1\,638^2 = 2\,683\,044 \equiv 884 \pmod{2\,579},\\
&y^8 = (y^4)^2 = 884^2 = 781\,456 \equiv 19 \pmod{2\,579},\\
&y^{16} = (y^8)^2 = 19^2 = 361 \pmod{2\,579},\\
&y^{32} = (y^{16})^2 = 361^2 = 130\,321 \equiv 1\,371 \pmod{2\,579},\\
&y^{64} = (x^{32})^2 = 1\,371^2 = 1\,879\,641 \equiv 2\,129 \pmod{2\,579},\\
&y^{128} = (x^{64})^2 = 2\,129^2 = 4\,532\,641 \equiv 1\,338 \pmod{2\,579}.
\end{aligned}$$

Käyttämällä yllä olevia tuloksia Bob saa laskettua, että

$$\begin{aligned}
D(y) = y^{27} &= y^{1+2+4+8+16+32+128} = y^1 y^2 y^4 y^8 y^{16} y^{32} y^{128}\\
&\equiv \underbrace{955 \cdot 1\,638}_{1564290 \equiv 1416} \cdot \underbrace{884 \cdot 19 \cdot 361}_{6063356 \equiv 127} \cdot \underbrace{1\,371 \cdot 1\,338}_{1834398 \equiv 729}\\
&\equiv \underbrace{1\,416 \cdot 127 \cdot 729}_{131097528}\\
&\equiv 1800 \pmod{2\,579}.
\end{aligned}$$

Bob muodostaa alla olevan taulukon ja purkaa sen avulla Alicen lähettämän salakirjoituksen loppuun. Kaikki taulukon tulokset ovat laskettu modulo 2 579.

y	y^2	y^4	y^8	y^{16}	y^{32}	y^{64}	y^{128}
955	1 638	884	19	361	1 371	2 129	1 338
2 188	720	21	441	1 056	1 008	2 517	1 265
1 687	1 332	2 451	910	241	1 343	928	2 377
1 628	1 751	2 149	1 791	1 984	702	215	2 382
1 018	2 145	89	184	329	2 502	771	1 271

$$D(955) = 955^{191} \equiv 1\ 800 \pmod{2\ 579} \to sa$$
$$D(2\ 188) = 2\ 188^{191} \equiv 504 \pmod{2\ 579} \to fe$$
$$D(1\ 687) = 1\ 687^{191} \equiv 1\ 517 \pmod{2\ 579} \to pr$$
$$D(1\ 628) = 1\ 628^{191} \equiv 812 \pmod{2\ 579} \to im$$
$$D(1\ 018) = 1\ 018^{191} \equiv 423 \pmod{2\ 579} \to ex$$

Alice lähetti Bobille viestin ”safe primex”. Koska sana ”primex” ei tarkoita mitään, Bob poistaa viimeisen x-kirjaimen ja saa viestiksi ”safe prime”. △

Pohlig-Hellmanin salakirjoitus lyhyesti:

- p on alkuluku, $\mathcal{P} = \mathcal{C} = \mathbb{Z}_p$ ja $\mathcal{K} = \{k \in \mathbb{Z}_p \mid \text{syt}(k,\ p-1) = 1\}$
- $E(x) = x^k \pmod{p}$, missä $x \in \mathcal{P}$ ja $k \in \mathcal{K}$
- $D(y) = y^{k^{-1}} \pmod{p}$, missä $y \in \mathcal{C}$ ja $k \in \mathcal{K}$

4 Salakirjoitusmenetelmien murtaminen

Edellisessä luvussa käytiin läpi historiallisesti merkittäviä salakirjoitusmenetelmiä, joilla Alice ja Bob voivat salata yksityiset viestinsä. Salausmenetelmää valitessa heidän kannattaa kuitenkin aina pohtia, kuinka helppo Even on murtaa menetelmä. Salausmenetelmän turvallisuutta ei takaa, että sen murtamiseen kuluisi useita vuosikymmeniä, sillä Eve on valmis käyttämään rajattomasti aikaansa tekemiinsä hyökkäyksiin. Välillä salauksen murtaminen voi tuntua Evestä jopa mahdottomalta, mutta hän ei silti anna periksi.

Jo Caesarin aikana pohdittiin, miten salakirjoitusmenetelmät voidaan muokata sellaisiksi, että vihollinen ei saa niitä murretuksi. Yksi merkittävä kysymys oli, onko salakirjoituksen turvallisuuden kannalta merkitystä, tietääkö hyökkääjä, millä menetelmällä salakirjoitus on laadittu? Esimerkiksi murtovaras näkee heti, millainen lukko ovessa on. Varkaalla ei ole kuitenkaan lukkoon sopivaa avainta ja näin ollen hän joutuu tekemään töitä murtaakseen oven.

Edeltävään kysymykseen ei saatu pätevää vastausta pitkiin aikoihin, kunnes vuonna 1883 hollantilainen Jean-Guillaume-Hubert-Victor-Frangois-Alexandre- Auguste Kerckhoff von Niuewenhof, lyhyemmin Auguste Kerchoff, esitti oman näkökulmansa asiaan [5]. Kerchoff päätyi ajatukseen, että salausmenetelmän turvallisuus ei voi pohjautua ajatukseen, että käytettävä menetelmä pysyisi salassa. Hänen mielestään salausmenetelmän toimintaperiaate ja siihen liittyvät yksityiskohdat tulisi olla kaikkien tiedettävissä lukuun ottamatta käytössä olevaa avainta. Tämä ajatus tunnetaan nykyisin *Kerchoffin periaatteena*. Ideana on, että kuka tahansa voi yrittää murtaa salauksen, mutta jos kukaan ei pysty siihen järkevässä ajassa, salausmenetelmää voidaan pitää melko luotettavana.

Useat salakirjoitusmenetelmät voidaan murtaa *raa'alla voimalla* (brute force). Raaka voima -menetelmässä salakirjoitus yritetään murtaa arvaamalla salaukseen

käytetty avain. Menetelmään soveltuvia avaimia kokeillaan siihen asti kunnes salakirjoituksesta saadaan selkokielinen viesti. Monissa salausmenetelmissä mahdollisten avainten määrä on kuitenkin niin iso, että universumin arvioitu elinaika, 10^{13} vuotta, ei riittäisi Evelle, vaikka hän kävisi läpi sekunnissa miljoona eri avainvaihtoehtoa.

Seuraavissa alaluvuissa esitellään tapoja, joita hyödyntämällä on murrettu symmetrisiä salausmenetelmiä. Murtamisien lähtökohtina ovat Kerckhoffin periaate ja oletus, että Eve on saanut käsiinsä pätkän salakirjoitetusta viestistä. Caesarin, affiini ja kertolaskusalakirjoitusmenetelmät voidaan murtaa kaikki samalla menetelmällä. Muissa edellä käydyissä salausmenetelmissä on taas omat heikkoutensa, joiden tietäminen auttaa salauksen murtamisessa. Päälähteinä on käytetty seuraavia lähteitä [4], [5], [8] ja [10].

4.1 Monoaakkosellisten korvausmenetelmien murtaminen

Kuten aiemmin nähtiin niin Caesarin salakirjoituksessa avain k kuuluu joukkoon $\mathbb{Z}_{26}$, jolloin mahdollisia avaimia on yhteensä 26 kappaletta. Luku 26 on melko pieni, eikä Eveltä menisi kovinkaan kauan tutkia, mitä lukua luvuista 0, 1, 2, ..., 26 Alice ja Bob käyttävät avaimenaan. Tästä syystä Caesarin salakirjoitus on melko helppo murtaa. Muinaisen Rooman aikana Caesarin käyttämä salakirjoitus oli kuitenkin melko turvallinen. Vaikka joku saikin käsiinsä hänen salakirjoitetun viestinsä, hän tuskin osasi murtaa sitä, sillä monet eivät osanneet edes lukea.

Kertolaskuun perustuvassa salakirjoituksessa on sama heikkous kuin Caesarin menetelmässä, ja itse asiassa se on vielä heikompi. Kertolaskusalakirjoituksessa on ainoastaan 12 mahdollista avainvaihtoehtoa, joten Evellä ei menisi kauaa käydessään ne läpi.

Affiini salakirjoitus on jo huomattavasti työläämpi murtaa raa'alla voimalla, sillä sen avain on lukupari $(m,\ k)$. Luku m voidaan valita 12 eri tavalla ja luku k 26 tavalla. Mahdollisia avaimeksi soveltuvia lukupareja on tällöin yhteensä $12 \cdot 26 = 312$. Lukuna 312 ei kuitenkaan ole niin iso, että Eve ei pystyisi kokeilemaan jokaista avainvaihtoehdon yksitellen läpi, vaikka hänellä olisi käytössään vain kynä ja paperi.

Raa'alla voimalla salauksien murtaminen on yksinkertaista, mutta se voi olla myös todella työlästä. Tämän vuoksi monet ovat yrittäneet keksiä vaivattomampia tapoja salausten murtamisiin. 800-luvulla arabialainen Abu Yusuf Yaqub ibn Ishaq al-Kindi teki merkittävän huomion yksinkertaisten korvausmenetelmien kryptoanalyysiä ajatellen. Jokaisessa kielessä osaa kirjaimista käytetään sanoissa useammin kuin toisia. Esimerkiksi suomen kielessä yleisimmät kirjaimet ovat a (11,62 %) ja i (10,71 %) ja harvinaisimmat å ja q, joita molempia esiintyy alle 0,01 % [7]. Englannin kielessä yleisimmät kirjaimet ovat taas e (12,70 %) ja t (9,06 %) ja harvinaisimmat q (0,09 %) ja z (0,07 %) [10]. Taulukossa 2 on esitetty vielä erikseen jokaisen kirjaimen esiintymistodennäköisyys englannin kielessä.

Ajatellaan nyt, että Eve on saanut käsiinsä osan Alicen ja Bobin toisilleen lähettämistä affiinilla salakirjoitusmenetelmällä salatuista viesteistä. Jos näissä viesteissä esiintyy muita kirjaimia enemmän esimerkiksi kirjainta H ja I, niin Eve voi tehdä arvauksen, että salakirjoituksessa esiintyvä H vastaa selkotekstin kirjainta e ja vastaavasti I vastaisi selkotekstin kirjainta t. Arvauksen pohjalta Eve saa muodos-

Taulukko 2: Englannin kielen kirjainten frekvenssit [10].

Kirjain	%	Kirjain	%	Kirjain	%	Kirjain	%
E	12,70	**H**	6,09	**W**	2,36	**K**	0,77
T	9,06	**R**	5,99	**F**	2,23	**J**	0,15
A	8,17	**D**	4,25	**G**	2,01	**X**	0,15
O	7,51	**L**	4,03	**Y**	1,97	**Q**	0,09
I	6,97	**C**	2,78	**P**	1,93	**Z**	0,07
N	6,75	**U**	2,76	**B**	1,49		
S	6,33	**M**	2,41	**V**	0,98		

tettua yhtälöparin, jonka ratkaisemalla hän saa tietoonsa mahdollisen avainparin (m,k). Arvaus ei aina johda oikeaan avaimeen, mutta tekemällä aina uuden arvauksen, Eve saa varmasti salakirjoituksesta selkokielisen viestin. Tätä kryptoanalyysissä käytettyä menetelmää kutsutaan *kirjainten frekvenssianalyysiksi*, ja sitä voidaan hyödyntää kaikissa yksinkertaisissa monoaakkosellisissa korvausmenetelmissä.

Esimerkki 4.1. Eve löysi julkisessa viestintäkanavassa liikkuneiden viestien joukosta Alicen Bobille lähettämän viestin "JCFWHF OHHEJS VICFIH WHJBPB JSVEHI IHUOUH NPHSTR FSFERB JBIZAH IHULJS HZPUXH RDCFIB CZPEAD HAZSZD". Eve tietää, että Alice on käyttänyt salaukseen affiinia salakirjoitusmenetelmää, mutta hän ei tietenkään tiedä, mikä avain $(m,\ k)$ Alicella on ollut käytössä. Eve yrittää murtaa salauksen kirjainten frekvenssianalyysillä.

Ensimmäiseksi Eve laskee salakirjoituksessa esiintyvien kirjainten frekvenssit ja listaa ne taulukkoon.

kirjain	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**
frekvenssi	3	5	4	3	4	6	0	14	7	6	0	1	0
kirjain	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
frekvenssi	1	2	4	0	3	6	1	4	2	2	1	0	5

Hän näkee taulukosta, että salakirjoituksessa esiintyy eniten kirjaimia H (14 kpl) ja I (7 kpl). Eve tekee tämän pohjalta arvauksen, että e (=5) on salautunut kirjaimeksi H (=7) ja kirjain t (=19) kirjaimeksi I (=8). Eve sijoittaa arvauksensa affiinissa menetelmässä käytettävään salausfunktioon ja saa yhtälöparin

$$\begin{cases} 4m + k = 7 \pmod{26} \\ 19m + k = 8 \pmod{26}. \end{cases}$$

Vähentämällä ylemmän yhtälön alemmasta Eve saa yhtälön

$$15m = 1 \pmod{26}$$

Kertomalla nyt saatu yhtälö puolittain luvun 15 käänteisluvulla modulo 26 saadaan ratkaistua vakion m. Eve etsii tarvitsemansa käänteisluvun Eukleideen algoritmin

ja Bézout'n yhtälön avulla.

Eukleideen algoritmi:	Bézout'n yhtälö:
$26 = 1 \cdot 15 + 11$	$1 = 4 - 3$
$15 = 1 \cdot 11 + 4$	$= 4 - (11 - 2 \cdot 4)$
$11 = 2 \cdot 4 + 3$	$= 3 \cdot 4 - 11$
$4 = 1 \cdot 3 + 1$	$= 3(15 - 11) - 11$
$3 = 3 \cdot 1 + 0$	$= 3 \cdot 15 - 4 \cdot 11$
	$= 3 \cdot 15 - 4(26 - 15)$
	$= 7 \cdot 15 - 4 \cdot 26$

Eve saa luvun 15 käänteisluvuksi 7, jolla hän kertoo aiemmin saamansa yhtälön puolittain.

$$\begin{aligned} & 15m = 1 \pmod{26} && \| \cdot 7 \\ \Rightarrow \quad & m = 7 \pmod{26} \end{aligned}$$

Vakio k saadaan ratkaistua sijoittamalla $m = 7$ alkuperäisen yhtälöparin ylempään yhtälöön.

$$\begin{aligned} & 4 \cdot 7 + k = 7 \pmod{26} && \| - 2 \\ \Rightarrow \quad & k = 5 \pmod{26}. \end{aligned}$$

Eve vähensi yhtälön molemmilta puolilta luvun 2, koska $4 \cdot 7 = 28 \equiv 2$ modulo 26. Eve on saanut ratkaistua arvauksensa pohjalta avaimeksi (7, 5).

Kun on saanut ratkaistua avaimen arvauksen pohjalta, kannattaa varmistaa, onko avain oikea, purkamalla sen avulla salakirjoitetua viestiä. Jos puretusta viestistä tulee järkevä, voi olla melko varma, että avain on sama, jota Alice ja Bob ovat käyttäneet. Affiini salakirjoitus puretaan funktiolla $D(y) = m^{-1}(y-k) \pmod{26}$, missä $y \in \mathcal{C}$. Aiemmin Eve laski, että luvun 15 käänteisluku on 7 modulo 26, jolloin siis $m^{-1} = 15$. Nyt Eve voi yrittää purkaa salakirjoituksen.

Salakirjoitus	Salakirjoituksen kirjainta vastaava numero y	$D(y) = 15(y-5) \pmod{26}$	Selkoteksti
J	9	15(9-5)=60, 60≡8 (mod 26)	i
C	2	15(2-5)=-45, -45≡7 (mod 26)	h
F	5	15(5-5)=0	a
W	22	15(22-5)=255, 255≡21 (mod 26)	v
H	7	15(7-5)=30, 30≡4 (mod 26)	e
F	5	15(5-5)=0	a

Ainakin alku näyttää Even kannalta lupaavalta, "I have a", joten Eve jatkaa purkamista. Koko viestiksi avaimella (7, 5) hän saa "I have a feeling that Eve is using letter frequency analysis to determine our key. What should we do now?" △

Sanojen frekvenssianalyysiä voidaan käyttää affiinien korvausmenetelmien murtamiseen. Sen vaiheet ovat:

- Lasketaan salakirjoituksessa esiintyvien kirjainten frekvenssit
- Verrataan saatuja frekvenssejä englannin kieleen ja tehdään arvaus, mitkä selkotekstin kirjaimet vastaisivat parhaiten salakirjoituksessa esiintyneitä kirjaimia
- Muodostetaan arvauksen pohjalta yhtälöpari, jonka ratkaisu on mahdollinen avain
- Avaimen oikeellisuudesta voi varmistua purkamalla salakirjoitus saatua avainta käyttäen. Jos purettu viesti on selkokielinen, oikea avain on luultavimmin löytynyt. Jos taas purettu viesti vaikuttaa sekavalta, tehdään uusi arvaus kirjainten vastaavuuksista.

Frekvenssianalyysin ehdoton heikkous on, että sen tehokkaaseen toimivuuteen tarvitaan melko pitkä salakirjoitus tai useampi salakirjoitettu viesti, joiden pohjalta kirjainten frekvenssit lasketaan. Jos Eve ei saa käsiinsä tarpeeksi pitkää pätkää salakirjoitettua viestiä, salakirjoituksessa esiintyvien kirjainten frekvenssit eivät välttämättä vastaa kielelle ominaisia suhteellisia frekvenssejä. Alice ja Bob voivat näin ollen lisätä viestien turvallisuutta lähettämällä ainoastaan hyvin lyhyitä viestejä ja vaihtamalla avainta säännöllisesti. Tällöin Eve ei pysty keräämään tarpeeksi hyödyllistä dataa frekvenssianalyysia varten, jolloin Even ainoaksi vaihtoehdoksi jää raaka voima.

4.2 Vigenéren salakirjoituksen murtaminen

Vigenéren salakirjoitusta ei ole yhtä helppo murtaa kuin monoaakkosellisia korvausmenetelmiä, sillä samaa selkotekstin kirjainta voi vastata useampi eri salakirjoituksen kirjain, minkä takia pelkästään kirjainten frekvenssianalyysilla ei pysty suoraan murtamaan Vigenéren salakirjoitusta.

Vigenéren salakirjoitus esiteltiin vuonna 1586, mutta kului lähes 300 vuotta ennen kuin julkaistiin ensimmäinen menetelmä sen murtamiseen. Vuonna 1864 entinen Preussin jalkaväen upseeri Friedrich Wilhelm Kasiski julkaisi oivalluksensa, joka tunnetaan nykyään *Kasiskin menetelmänä*. Kasiski ei ollut välttämättä ensimmäinen, joka keksi, miten Vigenéren salakirjoitus murretaan, mutta hän oli ensimmäinen, joka julkaisi sen ja teki näin historiaa [11]. Parhaimman tuloksen Kasiskin menetelmä antaa silloin, kun salauksessa on käytetty suhteellisen lyhyttä avainsanaa ja Evellä on runsaasti salakirjoitettua tekstiä.

Kirjainten frekvenssianalyysi perustui kielessä usein esiintyviin kirjaimiin. Kasiskin menetelmässä ei olla niinkään kiinnostuneita yksittäisistä kirjaimista vaan kielessä usein esiintyvistä kolmen kirjaimen pätkistä. Englannin kielessä esiintyy muita useammin esimerkiksi kolmikot ”the”, ”-ing”, ”and” ja ”for”. Kun salakirjoitettua tekstiä on tarpeeksi, on melko todennäköistä, että tällaiset pätkät salautuvat jossain kohtaa samalla tavalla. Näin tapahtuessa kolmikkojen väliin jäävän tekstin pituus on avaimen pituuden monikerta. Seuraava esimerkki voi selventää tilannetta.

Esimerkki 4.2. Viesti ”Make the guess that the...” salataan Vigenéren salakirjoitusmenetelmällä avainsanan ollessa ”key”.

m	a	k	e	**t**	**h**	**e**	g	u	e	s	s	t	h	a	t	**t**	**h**	**e**	...
k	e	y	k	e	y	k	e	y	k	e	y	k	e	y	k	e	y	k	...
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
W	E	I	O	**X**	**F**	**O**	K	S	O	W	Q	D	L	Y	D	**X**	**F**	**O**	...

Selkotekstissä esiintyy kahdessa kohtaa kolmikko ”the”, joka on molemmilla kerroilla salautunut kolmikoksi ”XFO”, joiden välillä on yhdeksän kirjainta. Avainsanan pituus oli kolme, ja yhdeksän on luvun kolme monikerta. △

Kun salakirjoituksesta löytyy useampi sama kolmen kirjaimen pätkä, voidaan avaimen pituus selvittää. Tehdään oletus, että nämä kolmikot vastaavat selkotekstissä samoja kirjaimia. Tällöin kolmikkojen väliin jäävät etäisyydet ovat kaikki avainsanan pituuden monikertoja, ja laskemalla näiden etäisyyksien yhteiset tekijät, voi tehdä arvauksen avainsanan pituudesta.

Kun Eve on saanut avaimen mahdollisen pituuden selville, hän voi jakaa tekstin yhtä moneen osioon kuin avaimessa on kirjaimia. Voidaan ajatella, että jokainen näistä osioista on salattu Caesarin salakirjoitusmenetelmällä eri avainta käyttäen, jolloin jokaiseen osioon voi yksitellen hyödyntään kirjainten frekvenssianalyysiä. Lopulta Eve saa mitä luultavimmin selville Alicen ja Bobin käyttämän avainsanan. Kasiskin menetelmän ymmärtäminen on helpompaa esimerkin avulla.

Esimerkki 4.3. Eve haluaa tietää, mitä Alice on kirjoittanut Bobille alla olevaan viestiin. Hän tietää, että Alice on käyttänyt salaukseen Vigenéren salausta, mutta muuta hän ei sitten tiedäkään. Eve yrittää murtaa salauksen Kasiskin menetelmällä.

Tutkittuaan hetken viestiä Eve huomaa, että viestissä on kolme kertaa peräkkäin kirjaimet E, I ja M (muut kolme kirjaimiset pätkät esiintyivät joko kerran tai kaksi), ja ne esiintyvät kohdissa 57, 67 ja 122.

Alicen Bobille lähettämä viesti:

IHRDKV EAMTTC VXJTRU IUQERV VWEHVD GENSCQ LRKKEH
RZUDFN ZPDTRD GCEIMJ PSNKJX EIMFPG RVGGAY AQAUGQ
QCHBEM PSVAMX DVLKIH RPCHTB JGHMNZ VTRGPC CEIM

Eve tekee nyt arvauksen, että EIM on saman kolmikirjaimisen selkotekstin pätkän salakirjoitus. Jos Even oletus on oikein, pätkien välimatkan on jaollinen avainsanan pituudella. Eve laskee, että kirjainyhdistelmien EIM väliin jää 67-57=10 ja 122-67=55 kirjainta. Lukujen 10 ja 55 suurin yhteinen tekijä on 5, joten Eve kokeilee avainsanan pituudeksi lukua 5, jolloin avainsanaa voidaan merkitä $k = (k_1,\ k_2,\ k_3,\ k_4,\ k_5)$.

Kun Evellä on arvaus avainsanan pituudesta, hän jakaa salakirjoituksen viiteen eri osioon, joissa salaukseen on käytetty samaa avainsanan kirjainta vastaavaa lukua k_i, missä $i = 1,\ 2,\ 3,\ 4,\ 5$. Siis ensimmäiseen osioon kuuluisivat kirjaimet I, V, T, T..., jotka ovat salattu avaimella k_1, toiseen osioon kuuluvat kirjaimet H, E, C, R,..., jotka ovat salattu avaimella k_2, ja niin edelleen. Jokainen viidestä osiosta on näin ollen salattu Caesarin menetelmällä käyttäen eri avainta k_i.

Kuten aiemmin todettiin Caesarin salakirjoitus voidaan murtaa kirjainten frekvenssianalyysilla, joten Eve tutkii jokaista viittä osiota erikseen. Eve laski, että ensimmäisessä osiossa esiintyi eniten C, P JA T-kirjaimia, joten hän tekee arvauksen, että englannin kielen yleisin kirjain e on tässä osiossa salautunut joksikin näistä kolmesta kirjaimesta. Tällöin hän saa arvauksestaan yhtälöt

$$\begin{aligned} 4 + k_1 = 2 \quad &\rightarrow k_1 = 24 \pmod{26}, \text{ tai} \\ 4 + k_1 = 15 \quad &\rightarrow k_1 = 11 \pmod{26}, \text{ tai} \\ 4 + k_1 = 19 \quad &\rightarrow k_1 = 15 \pmod{26} \end{aligned}$$

Eve toistaa saman päättelyn jokaiselle osiolle ja saa tulokseksi alla olevan taulukon.

	Osion kirjaimet	Osion yleisin kirjain	Avain
Osio 1 (k_1)	IVTTQWGQEDDC PXPGACPXIHHTC	C (3 kpl), P (3 kpl), T (3 kpl)	$k_1 = 24 \rightarrow$ y $k_1 = 11 \rightarrow$ l, $k_1 = 15 \rightarrow$ p
Osio 2 (k_2)	HECREEELHFTES EGAUHSDHTMRE	E (7 kpl),	$k_2 = 0 \rightarrow$ a
Osio 3 (k_3)	RAVURHNRRNRIN IRYGBVVRBNGI	R (7 kpl),	$k_3 = 13 \rightarrow$ n
Osio 4 (k_4)	DMXIVVSKZZDMK MVAQEALPJZPM	M (4 kpl),	$k_4 = 8 \rightarrow$ i
Osio 5 (k_5)	KTJUVDCKUPGJ JFGQQMMKCGVC	C (3 kpl), G (3 kpl), J (3 kpl), K (3 kpl)	$k_5 = 24 \rightarrow$ y, $k_5 = 2 \rightarrow$ c, $k_5 = 5 \rightarrow$ f, $k_5 = 6 \rightarrow$ g

Avainsana koostuu suurella todennäköisyydellä taulukon vaihtoehdoista, jolloin Evellä on avainsanaksi $3 \cdot 1 \cdot 1 \cdot 1 \cdot 3 = 9$ eri vaihtoehtoa. Vaikka avainsanaksi ei kannata valita selkeää sanaa, Eve kokeilee ensimmäisenä, olisiko Alice ja Bob valinneet avainsanaksi sanan ”panic”.

Salakirjoitus	Salakirjoituksen kirjainta vastaava numero y_i	$D(y_i) = y_i - k_1 \pmod{26}$	Selkoteksti
I	8	8-15=-7, -7≡19 (mod 26)	t
H	7	7-0=0	h
R	17	17-13=4	e
D	3	3-8=-5, -5≡21 (mod 26)	v
K	16	10-2=8	i
V	21	21-15=6	g
E	4	4-0=4	e
A	0	0-13=-13, -13≡13 (mod 26)	n
M	12	12-8=4	e
T	19	19-2=17	r

Even arvaus taisi osua oikeaan, ja kokonaisuudessa Alicen viesti oli:"The Vigenere cipher has been the unbreakable cipher so far. Not even Eve has achieved a general solution. How Kasiski did it? He has to be smarter than Eve." △

Vigenéren salakirjoitus voidaan murtaa noudattamalla seuraavia Kasiskin menetelmän vaiheita:

- Etsitään 3-kirjaimisia pätkiä, jotka toistuvat salakirjoituksessa
- Lasketaan, kuinka monta kirjainta edellä löydettyjen kirjainkolmikkojen väliin jää. Etsitään näiden välimatkojen yhteiset tekijät ja tehdään arvaus avaimen pituudesta.
- Jaetaan salakirjoitus yhtä moneen osioon kuin avaimessa on kirjaimia ja suoritetaan jokaiselle osiolle erikseen kirjainten frekvenssianalyysi
- Tehdään arvaus, mikä on kunkin osion avain, ja muodostetaan niistä arvaus avainsanaksi
- Tarkistetaan arvaus purkamalla salakirjoitusta tällä avaimella. Jos puretusta tekstistä ei muodostu selkokielistä tekstiä, tehdään avaimelle uusi arvaus.

Kasiskin analyysi ei johda aina haluttuun tulokseen. Jos Alice ja Bob lähettävät toisilleen suhteellisen lyhyitä viestejä, joiden salauksessa he käyttävät pitkää avainsanaa, Eve ei pysty hyödyntämään Kasiskin menetelmää ja ainoaksi vaihtoehdoksi jää jälleen raaka voima. Vigenéren salakirjoituksen murtamisessa raaka voima -menetelmä voi viedä kuitenkin aivan liian paljon aikaa, sillä Evellä ei ole mitään aavistusta avainsanan pituudesta. Joskus on jopa mahdollista, että Eve ei saa Vigenéren salakirjoitusta murrettua ollenkaan järkevässä ajassa.

4.3 Hillin salakirjoituksen murtaminen

Hillin salakirjoituksen murtaminen poikkeaa aiemmista tapauksista. Hillin salauksen murtamista hidastaa, että se on moniaakkosellinen. Moniaakkosellisuus ei ole kuitenkaan este murtamiselle. Suurempi ongelma Hillin salauksen murtamisessa on, että sille ei ole kehitetty yksinkertaista menetelmää avainsanan pituuden määrittämiseksi. Raa'an voiman käyttäminen on jälleen vaihtoehto, mutta Hillin menetelmässä matriisit tekevät laskuista työläämpiä.

Jos kuitenkin Kerchoffin periaatteen vastaisesti, Eve saa tietoonsa avaimen pituuden, hänellä on mahdollisuus murtaa salakirjoitus muullakin tavoin kuin raa'alla voimalla. Jos esimerkiksi avaimena on käytetty 2×2-matriisia, tiedetään, että salakirjoitus on saatu salaamalla kaksi selkotekstin kirjainta kerrallaan. Pilkkomalla nyt salakirjoitus kahden kirjaimen pätkiin voi tutkia, mitkä kirjainparit esiintyvät salakirjoituksessa useimmin. Vertaamalla saatuja frekvenssejä englannin kielessä useimmin esiintyviin kirjainpareihin (katso taulukko 3) voi tehdä arvauksen kirjainparien vastaavuuksista ja selvittää näin arvauksensa pohjalta, mitä avainta salaukseen on käytetty. Vastaavasti, jos salaus on tehty käyttämällä 3×3-matriisia, salakirjoitus

Taulukko 3: Englannin kielen kirjainparien frekvenssejä [8].

Kirjainpari	%	Kirjainpari	%	Kirjainpari	%
TH	6,3	**AR**	2,0	**HA**	1,7
IN	3,1	**EN**	2,0	**OU**	1,4
ER	2,7	**TI**	2,0	**IT**	1,4
RE	2,5	**TE**	1,9	**ES**	1,4
AN	2,2	**AT**	1,8	**ST**	1,4
HE	2,2	**ON**	1,8	**OR**	1,4

pilkottaisiin kolmen kirjaimen pituisiin pätkiin ja verrattaisiin niitä englannin kielessä useiten esiintyviin kolmen kirjaimen muodostamiin yhdistelmiin.

Esimerkki 4.4. Eve on jälleen löytänyt Alicen Bobille lähettämän salaviestin. Tällä kertaa Alice on salannut sen Hillin salakirjoitusmenetelmällä. Evelle kävi myös todella hyvä tuuri, kun hän sattumalta sai tietoonsa, että Alicen ja Bobin käyttämässä avaimessa on neljä kirjainta. Avain voidaan siis kirjoittaa muodossa $\begin{pmatrix} k_1 & k_2 \\ k_3 & k_4 \end{pmatrix}$, ja sillä salataan kaksi kirjainta kerrallaan. Eve jakaa salakirjoituksen kahden kirjaimen pareihin ja tutkii kirjainparien frekvenssejä. Hän huomaa, että kirjainparia CZ löytyy eniten (4 kpl) ja kirjainparia RT toisiksi eniten (3 kpl).

XL BZ KG EU BM HR ZT MG XT TM RT RX FT HY GU CZ YU
TM CH KO CC VG GW QB CZ DP GE MG CZ RT CE NO TX HL
VJ KO GK YN DP XC VG JO IA QB RT JO OZ CZ IK GM

Eve tutkii englannin kielen yleisempiä kirjainpareja taulukosta 3 ja tekee arvauksen, että salakirjoituksen kirjainpari CZ vastaa selkotekstin kirjaimia th ja RT vastaa kirjaimia in. Eve tekee arvauksestaan yhtälöt

$$(19\ 7)\begin{pmatrix} k_1 & k_2 \\ k_3 & k_4 \end{pmatrix} = (19\cdot k_1 + 7\cdot k_3 \quad 19\cdot k_2 + 7\cdot k_4) = (2\ 25) \pmod{26} \text{ ja}$$
$$(8\ 13)\begin{pmatrix} k_1 & k_2 \\ k_3 & k_4 \end{pmatrix} = (8\cdot k_1 + 13\cdot k_3 \quad 8\cdot k_2 + 13\cdot k_4) = (17\ 19) \pmod{26},$$

joista hän saa muodostettua yhtälöryhmän

$$\begin{cases} 19k_1 + 7k_3 = 2 \pmod{26} & \| \cdot 15 \\ 19k_2 + 7k_4 = 25 \pmod{26} & \| \cdot 15 \\ 8k_1 + 13k_3 = 17 \pmod{26} & \\ 8k_2 + 13k_4 = 19 \pmod{26} & \end{cases} \Rightarrow \begin{cases} 25k_1 + k_3 = 4 \pmod{26} \\ 25k_2 + k_4 = 11 \pmod{26} \\ 8k_1 + 13k_3 = 17 \pmod{26} \\ 8k_2 + 13k_4 = 19 \pmod{26}. \end{cases}$$

Eve kertoi yhtälöryhmän kaksi ensimmäistä yhtälöä luvulla 15, koska se on luvun 7 käänteisluku (ks. esimerkki 4.1). Tämän jälkeen hän vähensi ensimmäisen yhtälön luvulla 13 kerrottuna kolmannesta yhtälöstä ja vastaavasti neljännestä yhtälöstä hän vähensi toisen yhtälön luvulla 13 kerrottuna. Näin hän saa yhtälöparin, jonka

molemmat yhtälöt hän kertoo luvun 21 käänteisluvulla, joka on 5, ja saa

$$\begin{cases} 21k_1 = 17 \pmod{26} & \| \cdot 5 \\ 21k_2 = 6 \pmod{26} & \| \cdot 5 \end{cases} \Rightarrow \begin{cases} k_1 = 7 \pmod{26} \\ k_2 = 4 \pmod{26}. \end{cases}$$

Sijoittamalla edellä saadut arvot $k_1 = 7$ ja $k_2 = 4$ aiemmin saatuihin yhtälöihin $25k_1 + k_3 = 4 \pmod{26}$ ja $25k_2 + k_4 = 11 \pmod{26}$ Eve saa ratkaistua loppuun mahdollisen avaimen

$$\begin{cases} 25 \cdot 7 + k_3 = 4 \pmod{26} & \| -19 \\ 25 \cdot 4 + k_4 = 11 \pmod{26} & \| -22 \end{cases} \Rightarrow \begin{cases} k_3 = 11 \pmod{26} \\ k_4 = 15 \pmod{26} \end{cases}$$

Mahdollinen avain on $A = \begin{pmatrix} k_1 & k_2 \\ k_3 & k_4 \end{pmatrix} = \begin{pmatrix} 7 & 4 \\ 11 & 15 \end{pmatrix}$, jota vastaa sana ”help”. Eve tarkistaa vielä, onko sana ”help” oikea avain. Hän laskee ensimmäisenä avainmatriisille käänteismatriisin

$$\begin{aligned} A^{-1} &= (k_1k_4 - k_2k_3)^{-1} \begin{pmatrix} k_4 & -k_2 \\ -k_3 & k_1 \end{pmatrix} \pmod{26} \\ A^{-1} &= (7 \cdot 15 - 4 \cdot 11)^{-1} \begin{pmatrix} 15 & -4 \\ -11 & 7 \end{pmatrix} \\ &= 9^{-1} \begin{pmatrix} 15 & 22 \\ 15 & 7 \end{pmatrix} \quad \| 9^{-1} = 3, \text{ sillä } 3 \cdot 9 \equiv 1 \pmod{26} \\ &= 3 \begin{pmatrix} 15 & 22 \\ 15 & 7 \end{pmatrix} \\ &= \begin{pmatrix} 19 & 14 \\ 19 & 21 \end{pmatrix} \pmod{26}, \end{aligned}$$

jonka avulla hän purkaa salakirjoitusta

$$D(23,\ 11) = (23\ 11) \begin{pmatrix} 19 & 14 \\ 19 & 21 \end{pmatrix} = (23 \cdot 19 + 11 \cdot 19 \quad 23 \cdot 14 + 11 \cdot 21) = (646\ 553),$$

$$(646\ 553) \equiv (22\ 7) \pmod{26} \rightarrow \text{wh}$$

$$D(1,\ 25) = (1\ 25) \begin{pmatrix} 19 & 14 \\ 19 & 21 \end{pmatrix} = (1 \cdot 19 + 25 \cdot 19 \quad 1 \cdot 14 + 25 \cdot 21) = (494\ 539),$$

$$(494\ 539) \equiv (0\ 19) \pmod{26} \rightarrow \text{at}$$

$$D(10,\ 6) = (10\ 6) \begin{pmatrix} 19 & 14 \\ 19 & 21 \end{pmatrix} = (10 \cdot 19 + 6 \cdot 19 \quad 10 \cdot 14 + 6 \cdot 21) = (304\ 266),$$

$$(304\ 266) \equiv (18\ 6) \pmod{26} \rightarrow \text{sg}.$$

Ainakin salakirjoituksen alku näyttää purkautuvan selkokieliseksi tekstiksi avaimella ”help”, joten Eve on lähes varma, että hän löysi oikean avaimen. Alicen viesti oli kokonaisuudessaan: ”What’s going on? Eve is trying to break the cryptosystem all the time. I think, it is very possible, that Eve will invent the key.” △

Hillin salakirjoitus, jossa avaimena on käytetty neljäkirjaimista sanaa voidaan murtaa seuraavalla tavalla:

- Jaetaan salakirjoitus kahden kirjaimen pareihin ja lasketaan kirjainparien frekvenssit
- Verrataan saatuja frekvenssejä englannin kieleen ja tehdään arvaus, mitä selkotekstin paria yleisimmin salakirjoituksessa esiintyvät kirjainparit vastaavat
- Tehdään arvauksen pohjalta yhtälöryhmä, jonka ratkaisuna saadaan mahdollinen avain
- Tarkistetaan, onko saatu avain oikea, purkamalla salakirjoitus sitä käyttäen. Jos salakirjoituksesta purkautuu selkokielistä tekstiä, voidaan olettaa, että arvaus osui oikeaan ja ollaan löydetty oikea avain. Muussa tapauksessa tehdään uusi arvaus.

Hillin salakirjoituksen murtaminen on yleensä melko hankalaa, sillä avainsanan pituus on vaikea päätellä. Edellisessä esimerkissä Evelle kävi todella hyvä tuuri, kun hän oli saanut tietoonsa Alicen ja Bobin käyttämän avainsanan pituuden. Kechoffin periaate ei kuitenkaan ollut enää voimassa, sillä avaimesta oli paljastunut yksityiskohtia.

Salakirjoituksen turvallisuutta ajatellen Alicen ja Bobin kannattaa valita mahdollisimman pitkä avainsana, jolloin salauksen murtaminen vaikeutuu huomattavasti. On kuitenkin muistettava, että pitkä avainsana hidastaa myös Alicen ja Bobin työtä, sillä matriiseilla laskeminen on välillä melko kömpelöä. Lisäksi Even yritystä murtaa Hillin salakirjoitus hankaloittaa, että englannissa kirjainyhdistelmien suhteelliset frekvenssit ovat melko pieniä. On siis melko todennäköistä, että Even ensimmäiset veikkaukset eivät osu oikeaa ja salauksen murtaminen hidastuu.

4.4 Pohlig-Hellmanin salakirjoituksen murtaminen

Edellä ollaan nähty, että Eve on onnistunut murtamaan salakirjoitusmenetelmiä melko yksinkertaisillakin keinoilla. Hillin salakirjoitus tuotti Evelle ongelmia, mutta loppujen lopuksi myös sen salaus on niin heikko, että sitä ei kannata käyttää enää tänä päivänä. Luvussa 3.7 keskityttiin Pohlig-Hellmanin menetelmän tapaukseen, jossa salattiin aina kaksi kirjainta kerrallaan. Tämä erikoistapaus on mahdollista murtaa pelkällä raa'alla voimalla, mutta kukaan ei ole onnistunut vielä murtamaan Pohlig-Hellmanin salakirjoituksen yleistä versiota, kunhan sen käyttäjät ovat valinneet harkiten avaimensa.

Jos yrittää murtaa Pohlig-Hellmanin salakirjoitusta, päätyy nopeasti yhtälöön

$$x^k \equiv y \pmod{p},$$

josta pitäisi ratkaista salauseksponentti k. Tämän yhtälön ratkaisemista nimitetään toisinaan *diskreetin logaritmin ongelmaksi*, (DLP) [1]. Nimestä voi jo päätellä, että yhtälön ratkaiseminen ei ole kovinkaan suoraviivaista tai edes mahdollista. Pohlig

ja Hellman ratkaisivat itse diskreetin logaritmin ongelman, kun luku $p-1$ voidaan jakaa pieniin alkulukutekijöihin. Turvallisuuden kannalta alkuluku p tulisi siis valita siten, että luvulla $p-1$ olisi mahdollisimman suuret alkulukutekijät.

Jos alkuluku $p > 2$, niin $p-1$ on välttämättä jaollinen luvulla kaksi. Siis $p-1 = 2q$. Jos nyt luku q on alkuluku, niin luvulla $p-1$ ei ole muita alkulukutekijöitä kuin luvut 2 ja q. Jos luvuksi p on valittu tarpeeksi suuri alkuluku (noin 512-numeroinen) ja se voidaan kirjoittaa muodossa $2q+1$, jossa q on alkuluku, niin Pohlig-Hellmanin salakirjoitusmenetelmää voidaan pitää turvallisena. Ainakaan tähän mennessä kukaan ei ole vielä ratkaissut edellä mainittuja ehtoja täyttävää diskreetin logaritmin ongelmaa. Alkulukua p kutsutaan tällöin *turvalliseksi alkuluvuksi* (safe prime) ja alkuluku q on nimetty löytäjänsä mukaan *Sophie Germainin alkuluvuksi*.

Pohlig-Hellmanin salakirjoitus on siis pysynyt murtamattomana. Kun Alice ja Bob valitsevat salakirjoitukseen turvallisen alkuluvun ja eivät itse paljasta vahingossa avainta, he voivat viestitellä rauhassa Evestä välittämättä. Heidän on kuitenkin hyvä muistaa, että Eve voi yhä yllättää ja keksiä ratkaisun diskreetin logaritmin ongelmaan.

Viitteet

[1] G. H. Gadiyar, R. Padma: *The Discrete Logarithm Problem Over Prime Fields: The Safe Prime Case. The Smart Attack, Non-Canonical Lifts and Logarithmic Derivatives*. Czechoslovak Mathematical Journal, Vol. 68(4), pp.1115-1124, 2018

[2] L. S. Hill: *Cryptology in an Algebraic Alphabet*. The American Mathematical, Vol. 36(3), pp. 306-312, 1929.

[3] J. Hoffstein, J. Pipher, J. H. Silverman: *An Introduction to Mathematical Cryptography*. Springer, New York, 2008.

[4] J. Holden: *The Mathematics of Secrets: Cryptography from Caesar Ciphers to Digital Encryption*. Princeton University Press, Princeton, 2017.

[5] D. Kahn: *The Codebreakers: The story of Secret Writing*. The Macmillan Company, New York, 1973.

[6] S. C. Pohlig, M. E. Hellman: *An Improved Algorithm for Computing Logarithms over $GF(p)$ and Its Cryptographic Significance*. IEEE Transactions on Information Theory, Vol. 24(1), pp. 106-110, 1978.

[7] M. Pääkkönen: *A:sta Ö:hön: Suomen yleiskielen kirjaintilastoja*. Kielikello, Vol. 1, p. 3, 1991.

[8] A. Renvall: *Cryptography I*. Matematiikan ja tilastotieteen laitos, Turun yliopisto, Turku, 2012.

[9] A. Rondholz: *Crossing the Rubicon. A Histographical Study*. Mnemosyne, Vol. 62(3), pp. 432-450, 2009.

[10] S. Rubinstein-Salzedo: *Cryptography*. Springer, Cham, 2018.

[11] T. Schrödel: *Breaking Short Vigenére Ciphers*. Cryptologia, Vol. 32(4), pp. 334-347, 2008

[12] D. Underwood: *A guide to elementary number theory*. Mathematical Association of America, Washington, D.C., 2009.