Sissejuhatus

Selle konspekti esimene versioon kujutas endast osa Tartu Ülikoolis 1996. aasta kevadsemestril loetud arvuteooria[1] kursusest. Tänu sellele on ka paljud arvuteoreetilised tulemused võetud eelduseks, näiteks A1.1, A4.1, A4.1, A5.4, A7.1, A7.5, A9.1, A9.2, A9.6, A9.7 ja arutlus peatüki 9 alguses.

1997. aasta kevadel luges autor TTÜ-s juba krüptoloogia erikursust, mille raames sai konspekti algmeid kõvasti täiendatud. Võrgus ilmusid siis täiendatud kujul osad liht- ja jadašifritest ning RSA-st. Vastavalt TTÜ spetsiifikale ei olnud kursuse üldtoon aga eriti matemaatiline, erinevalt praegusest kursusest.

Seoses uue, 1997. sügisel loetava kursusega Tartu Ülikooli matemaatikateaduskonnas muutub konspekt oluliselt. Oluliseks teguriks ei ole siin ainuüksi ülikoolidevaheline erinevus, vaid eelkõige just autori suurem kogemus krüptoloogia vahendite praktilisel rakendamisel kommertssfääris, samuti vahepealse ajaga kogunenud teoreetiline pagas.

Käesolevas konspektis vaadeldakse krüptoloogiat kui teadust, mis tegeleb suurte andmevoogude korrektse käitlemise tagamisega tänapäevastes tehnilistes tingimustes (digitaalarvutite ning Interneti lai levik), sealjuures seletatakse lähemalt lahti sõna "korrektne" (ehk püstitatakse *turvaeesmärgid*, *information security objectives*) ning vaadeldatakse *krüptograafilisi primitiive* (*cryptographic tools*), mis on välja töötatud nende eesmärkide saavutamiseks, samuti meetodeid, kuidas erinevaid primitiive koondada ühtseks komplektiks, mille abil saab tänapäeval lahendada suhteliselt keerukaid krüptograafilisi probleeme.

Turvaeesmärkide ning krüptograafiliste primitiivide käsitlemisele on pühendatud konspekti esimene osa. Konspekti teises osas on toodud meetodid primitiivide kompleksеks käsitlemiseks. Konspekti viimane osa vaatab tulevikku, käsitledes kvantkrüptograafiat, mis põhineb tänapäevastest digitaalarvutitest fundamentaalselt erinevatel kvantarvutitel.

Ajaloost. Vajadust kaitsta oma andmeid erinevat liiki kurikaelade eest on inimesed tunnetanud juba ammu, kuid väga pika aja jooksul kasutati salakirjade vaid poliitilistel eesmärkidel: militaarias, diplomaatias jne. Krüptoloogia nö klassikalise perioodi heaks ülevaateks on [Kah67].

Arvutite ja sidesüsteemide plahvatuslik areng alates 60-ndatest aastatest on loonud ent täiesti uudse situatsiooni, kus ka eraisik vajab vahendeid, et kaitsta oma digitaalselt salvestatud pangaarvet ning armastuskirju kõikvõimalike erinevate kurikasutuste eest.

Kui 60-ndate lõpuni kasutatud krüptograafilised vahendid olid enamus realiseeritavad tahvli ja krihvli kaasabil (kuigi nende lahtimurdmine ei olnud alati sugugi lihtne, sakslaste *Enigma* šifri lahtimurdmine Teise Maailmasõja ajal poolakate ja inglaste poolt ([Kah67]) otsustas paljuski sõja tulemuse. Kuid — just Enigma koodi lahtimurdmiseks ehitati esimesed mehhaanilis-elektroonilised arvutid!), siis 1970-ndate alguses IBM'i teadlaste (Feistel jt) poolt loodud šiffer DES on juba loodud konkreetselt arvutitel realiseerimiseks (täpsemalt öeldes on DES üsna üheselt optimeeritud riistvaraliseks täitmiseks). DES kui 1977.

[1] `http://www.math.ut.ee/kursused/arvuteooria.html`

aastal USA-s mitteklassifitseeritud andmete kaitsmiseks vastuvõetud standard (esimene krüptostandard üldse) on ajaloos enim tuntud krüptograafiline mehhanism, olles siiamaani väga paljude finantsasutuste jaoks üle maailma *de dacto* standardiks.

Krüptoloogia kui teaduse alguseks loetakse 1976. aastal ilmunud Diffie ja Hellman'i artiklit "New Directions in Cryptography" ([DH76]), kus tutvustati *avaliku võtme krüptograafia* mõistet ning kirjeldati uut võtmevahetusmeetodit, mille turvalisus baseerub diskreetse logaritmi probleemi eeldataval "raskusel".

Avaliku võtme krüptograafia mõiste osutus revolutsiooniliseks. Ehkki originaalartiklis ei olnud autoritel välja tuua oma idee ühtegi praktilist realisatsiooni, oli idee väga lihtne ning tekitas teadusmaailmas suurt huvi ning aktiivsust. Juba 1978. aastal tõid Rivest, Shamir ja Adleman avalikuse ette esimese avaliku võtme krüptosüsteemi RSA ([RSA78]), mis baseerub teisel (arvatatavalt) raskel matemaatilisel probleemil: arvude faktoriseerimisel.

See, küllaltki ootamatu, klassikalise matemaatilise probleemi ('leida antud kordarvu mittetriviaalsed tegurid') *väga* praktiline rakendus taaselustas teadlaste huvi kiiremate faktoriseerimisalgoritmide leidmise vastu. 1980-ndatel aastatel tehti antud valdkonnas mitmed läbimurded, kuid ükski neist ei osutunud ohtlikuks RSA turvalisusele.

Üks kõige olulisemaid avaliku võtme krüptograafia rakendusi on digitaalne signatuur, mis võimaldab teha üheselt kindlaks teha andmeallika (andmeallika autentimine). Lisaks on välja töötatud palju erinevaid krüptograafilisi primitiive erinevate turvaeesmärkide saavutamiseks.

Uute avaliku võtme mehhanismide, eksisteerivate mehhanismide täiustuste ning turvatõestuste otsingud jätkuvad kogu aeg. Antud momendil on olemas suur hulk krüptostandardeid, andmeturvalisusega tegelevaid seadusi ning infrastruktuure.

Konspekti kõige esimese variandi koostamisel olid abiks Valdis Laan ja Ahto Buldas. Viimane on hea nõuga jätkuvalt abistanud ka järgnevate versioonide kirjutamisel.

Peatükk 1

Krüptograafia ülevaade

1.1 Andmeturve ja krüptograafia

Jättes vaatluse alt välja konkreetsed füüsilised kandjad tänapäevastes andmesidevõrkudes ning andmete kodeeringud, keskendume *informatsiooni* vahetamisele, mille käigus on vaja järgida (sõltuvalt situatsioonist ning konkreetsetest nõuetest) teatud hulka **turvaeesmärke**. Mõned tavalisemad turvaeesmärgid koos lühiseletustega on toodud joonisel 1.1.

Sajandite jooksul on välja töötatud hulk protseduure, mis adresseerivad turvaeesmärke juhul, kui tegu on füüsiliste dokumentidega (allkirjad, seifid, jne). Kuid päris sageli ei saa turvaeesmärke lahendada ainuüksi matemaatilisi algoritme kasutades, vaid tuleb toetuda seadustele ning organisatoorsetele vahenditele. Näiteks, kirjade privaatsust tagatakse margistatud ümbrikutega, mida levitab aktsepteeritud postiteenus. Ümbriku füüsiline turvalisus on praktilistel põhjustel limiteeritud; seetõttu on vaja ka seadusi, mis mõistavad süüdi ümbriku volitamata avaja. Samuti on ka digitaalse informatsiooni korral vaja aegajalt rakendada seaduse kaitset (digitaalsignatuuride seadus jms).

Füüsiliste andmekandjate korral on vahel turvalisuse saavutamiseks vaja peale informatsiooni enda turvata ka füüsiline kandja. Nii näiteks on paberraha turvatud magnetriba ja teiste vahenditega, mis raskendavad informatsiooni (antud juhul raha) kopeerimist.

Definitsioon 1.1.1 ***Digitaalne informatsioon*** *on informatsioon, mis ei ole seotud mingi konkreetse andmekandjaga.*

Digitaalse informatsiooni korral on andmete kopeerimine ja muutmine lihtne: andmetest võib teha miljoneid originaaliga absoluutselt samaseid koopiaid. Seega, on vaja vahendeid, mis võimaldavad turvata füüsilisest andmemeediumist sõltumatuid andmeid (turvamise all mõtleme siin olukorrast sõltuvate turvaeesmärkide tagamist).

Näide 1.1.2 (allkiri) *Üks esimesi asju, mida laps kirjutama õpib, on tema enda nimi. Ajapikku kujuneb inimesel välja kindel stiil, kuidas oma nime kirju-*

salastus (kondifentsiaalsus)	hoida informatsioon varjatud kõigi eest keda pole selle nägemiseks volitatud
andmeterviklikkus	teha kindlaks, et informatsiooni pole muudetud volitamata isikute poolt
olemi autentimine (identimine)	olemi (isiku, arvutiterminali, krediitkaardi) identiteedi kindlaksmääramine
andmeallika autentimine (sõnumiautentimine)	andmeallika kindlaksmääramine
signatuur	vahend, kuidas siduda informatsioon olemiga
volitamine	teisele olemile õiguse andmine olla miski või teha midagi
valideerimine	informatsiooni või ressursi kasutamiseks või manipuleerimiseks volituse hankimine
pääsukontroll	ressursidele pääsu lubamine ainult priviligeeritud olemitele
sertifitseerimine	informatsiooni signeerimine usaldatava olemi poolt
ajatempel	andmete loomise (või eksisteerimise) ajahetke salvestamine
tunnistajaks olemine	andmete loomise (või eksisteerimise) kontrollimine andmete loojast erineva olemi poolt
kviitung	kinnitus informatsiooni saamise kohta
konfirmatsioon	kinnitus teenuse saamise kohta
omandus	vahendid tagamaks olemile legaalne õigus kasutada või teisaldada antud ressurss
anonüümsus	mingisse protsessi kaasatud olemi identiteedi varjamine
salgamise vääramine	eelnenud tegevuste salgamise vääramine
tühistamine	sertifikaadi või volituse kehtetuks tunnistamine

Joonis 1.1: Mõned turvaeesmärgid

tada ehk allkiri, mida ühiskond hakkab teatud vanusest alates aktsepteerima kui inimese isiksuse täisväärtuslikku osa. Iga inimese allkiri on unikaalne, allkirja põhjal idendivad teised antud inimest unikaalselt. Digitaalse informatsiooni korral muutub ent allkirja mõiste, allkirjaks ei saa olla lihtsalt mingi unikaalne bitijada (kuna andmed on vabalt kopeeritavad!), vaid pigem mingi protseduur, mis sõltub signeeritavatest andmetest. Sellega seoses kõneleme me edaspidi digitaalsest signatuurist, *mitte allkirjast.*

Krüptoloogias kasutatakse analooge ka teiste "paberprotokollidega," krüptograafia eesmärgiks on välja töötada digitaalsed protseduurid turvaeesmärkide saavutamiseks, mis on vähemalt sama head kui nende "paberanaloogid." Samas ei ole olemas mingit garantiid, et kõikvõimalikke meid huvitavaid turvaeesmärke on võimalik digitaaljuhul üldse lahendada!

Definitsioon 1.1.3 ***Krüptograafia*** *on teadus, mis uurib erinevaid turvaeesmärke (vt joonis 1.1) tagavada aitavaid matemaatilisi tehnikaid ning meetodeid nende tehnikate ühendamiseks.*

Krüptograafilised eesmärgid

Joonisel 1.1 toodud turvaeesmärkidest on põhilised järgmised neli (neist kõik ülejäänud tuletatakse):

1. **Konfidentsiaalsus** (ehk *salastus*, *privaatsus*) — teenus, mille abil hoitakse andmed salajas volitamata olemite eest. Inglise keeles: *confidentiality*, *secrecy*, *privacy*.

2. **Andmeterviklikkus** (*data integrity*) on teenus, mis adresseerib andmete volitamata muutmist. Andmeterviklikkuse tagamiseks läheb vaja tehnikaid, mis võimaldavad tuvastada andmete manipuleerimist selleks volitamata olemite poolt. Andmete manipuleerimise all mõistetakse muuhulgas andmete kustutamist, lisamist ning asendamist.

3. **Autentimine** (*authentication*) on andmeterviklikkusega seotud teenus, mis rakendatakse nii olemitele kui andmetele. Kaks osapoolt, alustades suhtlemist, peavad end teineteisele identifitseerima. Üle kanali saadetud informatsioon peab olema autentitud (varustatud informatsiooniga andmeallika, saatmise aja jms kohta). Sellega seoses räägitakse kahest autentimise alamliigist: **olemi autentimisest** (*entity authentication*) ning **andmeallika autentimisest** (*data origin authentication*). Andmeallika autentimine tagab varjatud kujul ka andmeterviklikkuse (andmete modifitseerimisel muutub andmeallikas).

4. **Salgamise vääramine** (*non-repudiation*) on teenus, mille abil osutub olemil võimatuks salata eelnevalt võetud kohustusi või täidetud toiminguid. Kui olem salgab teatud kohustust ja tekivad vaidlused, peab olema vahend olukorra lahendamiseks. Näiteks võib üks olem volitada teist

olemit mingi ostu sooritamiseks ning pärast keelduda seda volitust tunnistamast. Sellistel juhtudel on vajalik olmanda, usaldatava, osapoole olemasolu.

Krüptograafia põhieesmärgiks on adekvaatselt adresseerida ülaltoodud nelja turvaeesmärki nii teoorias kui praktikas, ehk lihtsalt öeldes: krüptograafia on teadus selle kohta, kuidas vältida ja tuvastada pettust ja teisi pahategusid.

Järgnevates peatükkides vaatleme põhilise *krüptograafilisi primitiive*, mida kasutatakse andmeturbe tagamiseks. Joonisel 1.2 on loetletud mõningad krüptograafiliste primitiivid (täpsed definitsioonid tulevad hiljem) ning toodud nende vahelised seosed.

Primitiive hinnatakse erinevate kriteeriumite abil, mõned neist on toodud allpool.

1. **Turvatase** (*level of security*), mida mõõdetakse tavaliselt antud turvaeesmärgi nurjamiseks (kasutades parimaid antud momendil teatud meetodeid) vajalike operatsioonide arvuna.

2. **Funktsionaalsus**. Erinevate turvaeesmärkide saavutamiseks võib vaja olla primitiivide kombineerimist. Primitiivsete teatud põhiomadused määravad nende primitiivide efektiivsuse antud eesmärgi saavutamiseks.

3. **Töörežiim** (*methods of operation*). Erineval moel ja erinevate sisenditega rakendatud primitiivid pakuvad sageli väga erinevat funktsionaalsust.

4. **Töökiirus** (*performance*). Primitiivi efektiivsus kindlas töörežiimis (näiteks võib krüptimisalgoritmide kiirust mõõta krüptitud bittides sekundis).

5. **Realiseerimise lihtsus** (*ease of implementation*). Kui raske on primitiivi realiseerida antud vahendite olemasolu korral (näiteks riistvaras, tarkvaras).

Erinevate kriteeriumite suhteline tähtsus sõltub väga palju rakendusest ning kasutada olevatest ressursidest. Näiteks limiteeritud arvutusvõimsuse korral võib osutuda vajalikuks leppimine madalama turvatasemega, kuna kõrge turvataseme saavutamiseks antud arvutusvõimsusest ei piisa.

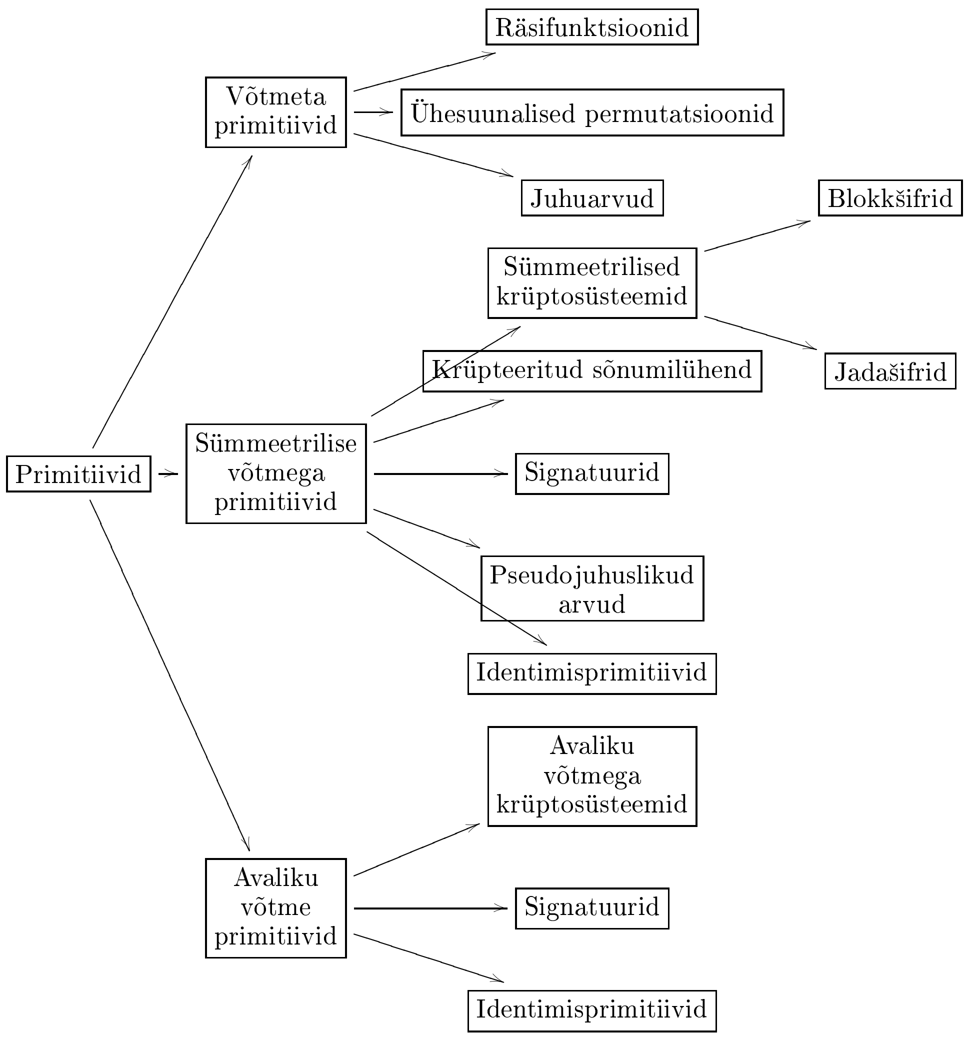
1.2 Baasterminoloogia ja -kontseptsioonid

Siin paragrahvis defineerime mõned mõisted, mis on olulised kogu edasises kursuses.

Definitsioon 1.2.1 *Sümboliga $\mathcal{A}$ märgime **tähestikku** (*alphabet*).*

*Sümboliga $\mathcal{P}$ märgime kõikvõimalike **avatekstide** (*plain text*) lõplikku hulka üle tähestiku $\mathcal{A}_P$ ning sümboliga $\mathcal{C}$ on kõikvõimalike **krüptogrammide** (*cipher text*) lõplikku hulka üle tähestiku $\mathcal{A}_C$, mis võib erineda tähestikust $\mathcal{A}_P$.*

*Sümboliga $\mathcal{K}$ märgime kõikvõimalike **võtmete** (*key*) lõplikku hulka.*



Joonis 1.2: Krüptograafilised primitiivid

Näide 1.2.2 *Tähestikuks võib võtta suvalise hulga, kus on rohkem kui üks element, kõige levinum näide on* $\mathcal{A} = \{0, 1\}$ *(binaarne tähestik). Tasub ent tähele panna, et enamus krüptograafilisi primitiive töötavad fikseeritud tähestikul!*

Avatekstide hulga näiteks võib olla eestikeelne tekst, arvutiprogrammide hulk jne.

Definitsioon 1.2.3 *Tähistagu* $\mathcal{E}$ ***krüptimisfunktsioonide*** *ning* $\mathcal{D}$ ***dekrüptimisfunktsioonide*** *hulka, kus igale võtmele* $K \in \mathcal{K}$ *vastab parajasti üks bijektsioon* $E_K \in \mathcal{E} : \mathcal{P} \to \mathcal{C}$ *ning parajasti üks bijektsioon* $D_K \in \mathcal{D} : \mathcal{D} \to \mathcal{C}$.

Funktsiooni E_K *rakendamist elemendile* $p \in \mathcal{P}$ *nimetatakse* ***krüptimiseks*** *(ka krüpteerimine,* encryption*), funktsiooni* D_K *rakendamist elemendile* $c \in \mathcal{C}$ *nimetatakse* ***dekrüptimiseks*** *(ka dekrüpteerimine,* decryption*).*

Definitsioon 1.2.4 ***Krüptosüsteem*** *(*encryption scheme*) ehk* ***šiffer*** *(*cipher*) koosneb viisikust* $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$*, nii et suvalise* $e \in \mathcal{K}$ *korral leidub unikaalne* $d \in \mathcal{K}$*, nii et* $D_d = E_e^{-1}$*, st* $D_d(E_e(p)) = p$ *suvalise* $p \in \mathcal{P}$ *korral. Sealjuures paari* (e, d) *nimetatakse* ***võtmepaariks****.*

1.2.1 Konfidentsiaalsuse saavutamine

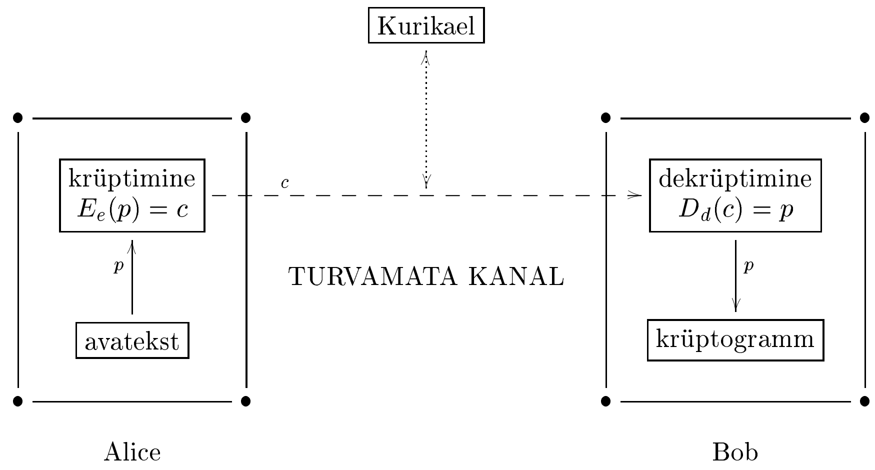
Kaks osapoolt Alice ja Bob valivad esmalt turvaliselt (või vahetavad turvaliselt) võtmepaari (e, d). Mingil suvalisel ajahetkel kui Alice soovib Bobile saata avateksti $p \in \mathcal{P}$, krüptib ta selle võtmega e ning edastab Bobile krüptogrammi $c := E_e(p)$. Krüptogrammi c kätte saamisel leiab Bob suuruse $D_d(c)$, mis on definitsiooni kohaselt võrdne suurusega p.

Märkus 1.2.5 *Siinkohas võidakse küsida, miks on võtmed üleüldse vajalikud (miks mitte valida lihtsalt krüptimisfunktsioon ja sellele vastav dekrüptimisfunktsioon?). Vastus: erinevate funktsioonide, mis on väga sarnased ning karakteriseeritud mingi lõpliku pikkusega võtme* K *abil, kasutamine tähendab, et konkreetse (de)krüptimisfunktsiooni avalikuks tulekul ei ole vaja ümber ehitada kogu krüptosüsteemi, vaid lihtsalt valida teine võti. Praktikas on mõistlik võtmeid (vastavaid funktsioone) vahetada sageli, et mitte anda kurikaeladele aega võtme lahti murdmiseks. Füüsiliseks analoogiks oleks siinkohal tavaline kombinatsioonidega lukk. Luku valmistab tehas ja see on kättesaadav kõigile (kel raha on), kuid avava kombinatsiooni valib omanik. Kui omanik kahtlustab kombinatsiooni avalikuks tulekut, saab ta seda lihtsalt muuta ilma füüsilist konstruktsiooni muutmata.*

Kahe osapoolega krüptitud suhtlus on skemaatiliselt kujutatud joonisel 1.3.

1.2.2 Osapooled

Definitsioon 1.2.6 ***Olem*** *(*entity*) ehk* ***osapool*** *(*party*) on keegi või miski, kes või mis saadab, võtab vastu ning manipuleerib informatsiooni (joonisel 1.3 on olemiteks Alice, Bob ning kurikael). Olemiks võib olla inimene, ATM jms.*



Joonis 1.3: Kahe osapoolega suhtlus, kasutades krüptimist

Saatja *(*sender*) on kahepoolses suhtluses informatsiooni volitatud edastaja (joonisel Alice).*

Vastuvõtja *(*receiver*) on kahepoolses suhtluses olem, kellele informatsioon mõeldud oli (joonisel Bob).*

Kurikael *(*adversary, enemy, attacker, opponent, tapper, eavesdropper, intruder, interloper*) on kahepoolses suhtluses olem, kes ei ole ei saatja ega vastuvõtja ning kes üritab nurjata andmeturbe teenust saatja ja vastuvõtja vahel. Kurikael üritab suhtluses sageli mängida saatja või vastuvõtja rolli. Kurikaela standardnimed on: Oscar, Carol, Malice.*

1.2.3 Kanal

Definitsioon 1.2.7 ***Kanal*** *(*channel*) on vahend edastamaks informatsiooni ühelt olemilt teisele.* ***Füüsiliselt turvaline kanal*** *(ehk lihtsalt* ***turvaline kanal****,* secure channel*) on kanal, mis ei ole kurikaelale füüsiliselt kättesaadav.* ***Turvamata kanal*** *on kanal, kus peale saatjate ja vastuvõtjate ka teised olemid saavad informatsiooni lugeda, kustutada, lisada või asendada.* ***Turvatud kanal*** *on kanal, kus kurikael ei saa informatsiooni lugeda, kustutada, lisada või asendada.*

Märkus 1.2.8 *Erinevus füüsiliselt turvalise kanali ja turvatud kanali vahel on selles, et viimane võib olla kas füüsiliselt turvaline või turvatud krüptograafiliste vahendite abil (krüptograafia tegeleb just viimast tüüpi kanalite ehitamisega). Eeldatakse, et teatud kanalid (isiklik kontakt, usaldatud saadik, spetsiaalne ühenduskanal) on füüsiliselt turvalised, kuid tasub meelde jätta, et 100% kindel ei saa ühegi kanali turvalisuses olla*[1].

[1] Üks kvantkrüptograafia olulisi saavutusi on füüsiliselt turvalise kanali olemasolu, kui ka-

1.2.4 Konfidentsiaalsus

Krüptograafia üks fundamentaalsemaid eeldusi on see, et krüptosüsteem ise (st viisik $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$) oleks avalik. Kui kaks osapoolt soovivad turvaliselt suhelda, kasutades antud krüptosüsteemi, hoiavad nad saladuses vaid konkreetset võtmepaari (e, d). On küll võimalik saavutada suuremat konfidentsiaalsust, hoides krüptimis- ja dekrüptimisfunktsioonid saladuses, kuid krüptosüsteemi turvalisus ei tohi baseeruda sel lähenemisel! Ajalugu on näidanud, et funktsioonide salajas hoidmine on väga raske.

Definitsioon 1.2.9 *Öeldakse, et krüptosüsteem on* ***lahti murtav*** *(*breakable*), kui kurikael saab võtmepaari (e, d) teadmata taastada avatekste süstemaatiliselt neile vastavatest krüptogrammidest mingi "sobiva" aja jooksul.*

Siinkohas "sobiv aeg" sõltub otseselt turvatava informatsiooni kasulikust eluajast. Nii näiteks, käsku "osta Ühispanga aktsiaid" on vaja varjata vaid loetud arv minuteid, riigisaladusi teinekord ent kuni universumi lõpuni.

Krüptosüsteemi saab murda **täieliku läbivaatamise meetodiga** (*exhaustive search*), mis seisneb antud krüptogrammi dekrüptimises kõikvõimalike erinevate võtmetega. Siit tuleb järgmine tähtis järeldus:

> võtmehulk $\mathcal{K}$ peab olema piisavalt suur, et täieliku läbivaatamise meetod ei oleks edukas.

See tähendab, et kogu inimkonna geniaalsusest jääb väheks, et ehitada kõrge turvatasemega krüptosüsteem, kui $|\mathcal{K}|$ on väike.

Krüptograafide eesmärgiks on disainida selliseid krüptosüsteeme, mille korral parim meetod krüptosüsteemi lahtimurdmiseks töötab samas ajas kui täieliku läbivaatamise meetod.

Kirjanduses tsiteeritakse sageli järgmisi Kerckhoff'i poolt 1883. aastal kirja pandud kriteeriume krüptosüsteemidele, mis on enamuses kasulikud ka tänapäeval:

1. Kui krüptosüsteem ei ole teoreetiliselt lahti murdmatu, peab ta olema murdmatu praktikas.

2. Krüptosüsteemi avalikuks tulek ei tohi osapooli ohustada.

3. Võti peab olema meelde jäetav ilma märkmeteta ning kergesti vahetatav.

4. Krüptogramm peab olema edastatav telegraafiga.

5. Krüptimisaparatuur peab olema kaasaskantav ning opereeritav üksikisiku poolt.

6. Krüptosüsteem peab olema lihtne, nõudmata pika nimekirja reeglite teadmist või muud võrreldavat mentaalset pingutust.

Kriteeriumi number 2 tuntakse ka kui lihtsalt **Kerckhoff'i kriteeriumit**.

---

sutada kvantmehhaanika vahendeid. "Tavamaailmas" tõestatavalt turvalisi kanaleid ent ei leidu.

1.2.5 Üldine andmeturve

Siiani oleme terminoloogia osas piirdunud vaid krüptimise ja dekrüptimisega. Andmeturve on ent märksa laiem ala, sisaldades peale konfidentsiaalsuse veel näiteks autentimist ja andmeterviklikkust. Anname siinkohas veel mõned üldised definitsioonid.

Definitsioon 1.2.10 ***Andmeturbe teenus*** *(*information security service*) on meetod, mis tagab andmeturbe mingit spetsiifilist aspekti. Näiteks, edastatud andmete terviklikkus on turvaeesmärk, selle eesmärgi tagamise meetod on aga andmeturbe teenus.*

Andmeturbe teenuse (mis sisaldab tihti muudki kui lihtsalt krüptimist) ***lahti murdmise*** *all mõeldakse antud teenuse eesmärgi nurjamist.*

Passiivne kurikael *on kurikael, kes suudab turvamata kanalilt andmeid lugeda, kuid mitte enamat.*

Aktiivne kurikael *on kurikael, kes suudab turvamata kanalil andmeid ka kustutada, muuta ja edastada.*

1.2.6 Krüptoloogia

Definitsioon 1.2.11 ***Krüptoanalüüs*** *(*cryptanalysis*) on teadus, mis tegeleb andmeturbe teenuste nurjamiseks mõeldud matemaatiliste tehnikate uurimisega.*

Krüptoloogia *(*cryptology*) on teadus, mis ühendab endas krüptograafia ja krüptoanalüüsi.*

Krüptosüsteem *on üldine termin, millega tähistatakse mingi andmeturbe teenuse pakkumiseks vajaminevate krüptograafiliste primitiivide hulka (kuigi tavaliselt kasutatakse mõistet vaid konfidentsiaalsusest rääkides).*

Järgnevates peatükkides teeme lähemat tutvust erinevate krüptograafiliste primitiividega.

Peatükk 2

Sümmeetrilised krüptosüsteemid

2.1 Blokkšifrid. Baasdefinitsioonid

Definitsioon 2.1.1 . ***Sümmeetriliseks krüptosüsteemiks*** *nimetatakse viisikut*

$$(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D}),$$

mis rahuldab definitsiooni 1.2.4 nõudeid kui leidub ökonoomne algoritm, mis iga võtmepaari esimesele elemendile e seab vastavusse teise elemendi d.

Blokkšiffer *(*block cipher*) on sümmeetriline krüptosüsteem, mille korral avatekst jaotatakse võrdpikkusega lõikudeks (blokkideks) $p_1, p_2, \ldots \in \mathcal{A}^t$ ning avatekst krüptitakse blokkhaaval.*

Fakt, et enamuses praktilistes sümmeetrilistes krüptosüsteemides $e = d$, seletab sõna "sümmeetriline" kasutamise antud kontekstis.

Konkreetse krüptosüsteemi järgimiseks tuleb täita järgnevad sammud: kõigepealt valivad Alice ja Bob juhusliku võtme $e \in \mathcal{K}$ (selleks *peab* olema turvaline kanal). Edasises oletagem, et Alice tahab Bob'ile edastada stringi

$$p := p_1 p_2 \ldots p_n,$$

kus $n \geq 1$ ning iga i korral $p_i \in \mathcal{P}$ (nagu on juba korduvalt rõhtutatud, on võti d ainuke element, mis hoitakse salajas. Kuna aga sümmeetriliste krüptosüsteemide korral on d kergesti leitav võtmest e, peab ka e salajane olema).

Peale krüptosüsteemi lepivad Alice ja Bob eelnevalt kokku ka **töörežiimis**, st viisis, kuidas sõltuvad teineteisest avateksti järjestikuste elementide šifreeringud. Alice šifreerib p'i ning saadab Bobile stringi

$$c := c_1 c_2 \ldots c_n.$$

Bob dešifreerib saadud krüptogrammi, kasutades sama krüptosüsteemi, töörežiimi ja võtit.

Toome siinkohas ära kahe lihtsama režiimi kirjeldused (dekrüptimisprotsess on kummalgi juhul triviaalselt tuletatav krüptimisreeglitest):

- Kõige lihtsam, **koodiraamatu režiim** (ECB, *electronic codebook*):

$$(\forall i) c_i := E_e(p_i)$$

- **Ahelrežiim** (CBC, *cipher block chaining mode*):

$$\begin{aligned} c_0 &:= E_e(p_0) \\ c_{i+1} &:= E_e(p_{i+1} \oplus c_i) \end{aligned}$$

(kus $\oplus$ tähistab bitikaupa välistavat võid).
Seega stringi p šifreering sõltub

1. kasutatavast krüptosüsteemist,
2. töörežiimist.
3. salajasest võtmest K.

Töörežiimide pikem võrdlus tuleb konspektis hiljem.

Blokkšifreid võrreldakse paragrahvis 1.1 toodud viie kriteeriumi järgi, ka sellest räägime natuke hiljem. Meelde tuleks jätta järgmised nõuded blokkšifritele.

1. *Turvatase.* Hea blokkšiffri lahti murdmiseks kulub väga kaua aega.
2. *Funktsionaalsus.* Blokkšifrid on eelkõige mõeldud konfidentsiaalsuse saavutamiseks, kuid erinevates töörežiimides sobivad ka näiteks autentimiseks. Head blokkšifrit on üldjuhul võimalik kasutada erinevate turvaeesmärkide saavutamiseks ning ta integreerub hõlpsasti teiste primitiividega.
3. *Töökiirus.* Tänapäeval on laialt kasutatavad piisavalt kõrge turvatasemega blokkšifrid küllaltki sarnase töökiirusega. Praktikas tähendab teinekord juba 10 protsendine kiiruse vahe aeglasema blokkšifri "halvaks" tunnistamist.
4. *Realiseerimise lihtsus.* "Juhuslik blokkšiffer" võtab enda alla eksponentsiaalse ruumi, hea blokkšiffer on efektiivselt realiseeritav riist- ja/või tarkvaras.

2.2 Lihtšifrid

Lihtšifreid jaotatakse kahte alamliiki: substitutsiooni- ning permutatsioonišifrid.

2.2.1 Substitutsioonišifrid

Sümboliga $\mathcal{S}_X$ tähistame hulga X kõigi substitutsioonide rühma.

Krüptosüsteem 2.2.1 (substitutsioonišiffer) .
Olgu $\mathcal{A}$ - tähestik, kus on q sümbolit ning $\mathcal{P} = \mathcal{C} = \mathcal{A}^t$. Olgu $\mathcal{K} = \mathcal{S}_{\mathcal{P}}$. Iga $e \in \mathcal{K}$ korral defineeritakse $E_e \in \mathcal{E}$ järgmiselt:

$$E_e(p) := e(p_1) \cdots e(p_t) = c_1 \cdots c_t = c$$

St, iga sümbol m_i asendatakse sümboliga $e(m_i)$. Dekrüptimine käib täpselt analoogiliselt. Olgu $d = e^{-1}$, siis

$$D_d(c) := d(c_1) \cdots d(c_t) = p_1 \cdots p_t = p$$

Nii üldise ülesandepüstituse korral $|\mathcal{K}| = q!$ (ei sõltu bloki pikkusest!) ning seega on brutaalse jõuga krüptoanalüüs raskendatud (piisavalt suure $|\mathcal{P}|$ korral).

Näide 2.2.2 *Edaspidistes näidetes vaatleme juhtu, kus $\mathcal{P} = \mathbf{Z}_{32}$. Defineerime kujutuse μ eesti tähtede hulkast hulka $\mathbf{Z}_{32}$ järgmiselt: $\mu($`A`$) = 0$, $\mu($`B`$) = 1$ jne (vt joonis 2.1).*

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Q	R	S	Š	Z	Ž	T	U	V	W	Õ	Ä	Ö	Ü	X	Y
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31

Joonis 2.1: Eesti tähed

Krüptosüsteem 2.2.3 (nihkešiffer) . *Olgu $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbf{Z}_{32}$ ja $e \in \mathcal{K}$, defineerime*

$$E_e(p) = p + e \pmod{32}$$

ning $d = e$,

$$D_d(c) = c - d \pmod{32}.$$

Erijuhul $e = 3$ nimetatakse seda krüptosüsteemi ka **Caesari šifriks** Caesari kui selle šifri väidetava esmakasutaja auks.

Näide 2.2.4 *Olgu $e = 6$, $x =$ MÄEKÜLAPIIMAMEES. Šifreerides saame ($\mu($M$) + 6 = \mu($S$)$ jne) krüptogrammiks stringi SBKQDRGŽOOSGSKKV (eesti keele liiasuse tõttu on küllalt raske leida pikemaid sõnu, mis nihkešifriga šifreeritult tähendust omaks). Antud krüptogrammi saamise järel rakendab Bob vastandprotseduuri: $\mu($S$) - 6 = \mu($M$)$, $\mu($B$) - 6 = \mu($Ä$)$ jne.*

Pangem tähele, et nihkešiffer ei ole turvaline, kuna teda saab täieliku läbivaatusmeetodit kasutades krüptoanalüüsida: tuleb läbi vaadata ainult $|\mathcal{A}|$ erineva võtmega dekrüptimisel saadud avateksti ning valida nende hulgast välja "mõistlikuna" näivad.

Harjutus 2.2.5 . *Leida täieliku läbivaatusmeetodiga krüptogrammile* ÄRÕRZÕRAWWÜ *vastav avatekst.*

Definitsioon 2.2.6 *Olgu* $n \in \mathbf{Z}^+$. ***Euleri funktsiooni*** ϕ *väärtuseks kohal* n *on* n*-ist väiksemate* n*-iga ühisteguriteta naturaalarvude arv.*

Krüptosüsteem 2.2.7 (afiinne šiffer(*affine cipher*)) *Olgu* $\mathcal{P} = \mathcal{C} = \mathbf{Z}_{32}$ *ning*

$$\mathcal{K} = \{(a, b) \in \mathbf{Z}_{32} \times \mathbf{Z}_{32} : \gcd(a, 32) = 1\}.$$

Iga $e = (a, b) \in \mathcal{K}$ *jaoks defineerime*

$$E_e(p) = ap + b \pmod{32} \text{ ning}$$
$$D_e(c) = a^{-1}(c - b) \pmod{32}.$$

Tingimus $\gcd(a, 32) = 1$ on vajalik, kuna vastasel juhul poleks E_e bijektsioon. Antud krüptosüsteemi korral $|\mathcal{K}| = 32 \cdot \phi(32) = 32 \cdot 16 = 512$. Kontrollime: $D_e(E_e(x)) = a^{-1}((ax + b) - b) \equiv x \pmod{32}$.

Anname järgnevalt algoritmi $a^{-1} \mod 32$ leidmiseks. Peatume sellel siinkohal lühidalt, kuna meil läheb seda algoritmi hiljem korduvalt vaja.

Algoritm 2.2.8 (Eukleidese algoritm) *Olgu* $\gcd(a, n) = 1$, *vaja leida* $a^{-1} \mod n$.

<u>funct</u> $inv(a, n) \equiv$
  $r_0 = n$; $r_1 := a$; $i := 0$; $t_0 := 0$; $t_1 := 1$;
  <u>while</u> *true* **<u>do</u>**
    $q_{i+1} := \lfloor r_i / r_{i+1} \rfloor$;
    $r_{i+2} := r_i - r_{i+1} q_{i+1}$;
    $t_{i+2} := t_i - q_{i+1} t_{i+1} \mod n$;
    <u>if</u> $r_{i+2} = 0$ **<u>then</u>** $inv := t_{i+1} \mod n$; **<u>exit</u>**; **<u>fi</u>**
    $i := i + 1$;
  <u>od</u>.

Harjutus 2.2.9 . *Tõestada, et* $a^{-1} = t_{i+1} \mod n$.

Toome tõestamata ära

Teoreem 2.2.10 (Lamé teoreem) *Kui Eukleidese algoritmis on* m *sammu, siis* $f_{m+2} \leq n$, *kus* f_i *tähistab* i*-ndat Fibonacci arvu (induktsiooniga,* $f_{m+2} = f_{m+1} + f_m \leq r_1 + r_2 \leq r_0 = n$*).*

Harjutus 2.2.11 *Tõestada Lamé teoreem.*

Kuna

$$f_i \approx \left(\frac{1+\sqrt{5}}{2}\right)^i ,$$

siis $m = O(\log n)$ ning Eukleidese algoritmi keerukus on $O((\log n)^3)$ (tegelikult natuke rohkem pingutades saaksime hinnangu $O((\log n)^2)$).

Harjutus 2.2.12 *Võttes $K = (3,1)$, šifreerida nimi* `OTTOKATZ`*. Sama võtme korral dešifreerida krüptogramm* `QNCKYGZBBD`*.*

Üldiselt, lihtsad substitutsioonišifrid üle väikeste blokipikkuste ei paku olulist konfidentsiaalsust ka siis, kui $\mathcal{K}$ on väga suur. Nimelt säilitavad substitutsioonišifrid avateksti tähtede sageduse, kõrge liiasusega tekstide krüptimisel (kõikidel loomulikel keeltel on *väga* kõrge liiasus!) osutuvad sellised šifrid kergesti murtavateks, kui analüüsida tagasihoidliku pikkusega krüptogrammi. Statistilist analüüsi pärsib liiasuse vähendamine, üks võimalus selleks on avateksti juhuslike elementide lisamine. Samas teeb see krüptogrammi pikemaks ning selliste šifrite korral (vaata järgnevat definitsiooni) on dekrüptimine raskem.

Krüptosüsteem 2.2.13 (homofooniline substitutsioonišiffer) *Igale sümbolile $a \in \mathcal{A}$ seatakse vastavusse pikkusega t stringide hulk $H(a)$ nii, et $a \neq b$ korral $H(a) \cup H(b) = \emptyset$. Homofooniline substitutsioonišiffer asendab iga avateksti sümboli a juhuslikult valitud stringiga hulgast $H(a)$. Pikkusega t stringi p dekrüptimiseks tuleb leida selline a, et $p \in H(a)$. Šifri võti koosneb hulkadest $H(a)$.*

Näide 2.2.14 *Olgu $\mathcal{A} = \{0,1\}$ ning suvalise $a \in \mathcal{A}$ korral $H(a) = \{0a, 1a\}$. Avatekst 01 krüptitaks sellisel juhul üheks järgmistest stringidest: 0010, 0011, 0110, 0111. Antud juhul vastab igale pikkusega 2 bitistringile täpselt üks $\mathcal{A}$ element.*

Eelpool toodud krüptosüsteeme nimetatakse **monoalfabeetilisteks**, kuna iga valitud võtme korral kujutub suvaline avateksti sümbol konstantseks sümboliks. **Polüalfabeetiline substitutsioonišiffer** on selline blokkšiffer, mille korral

1. Hulk $\mathcal{K}$ koosneb kõikvõimalikest permutatsioonijadadest $(e_1, \dots, e_t)$ pikkusega t, kus $e_i \in \mathcal{S}_{\mathcal{A}}$.

2. Sõnumi $p = p_1 \dots p_t$ krüptimisel võtmega $(e_1, \dots, e_t)$ saadakse krüptogramm $E_e(p) = e_1(p_1) \dots e_t(p_t) = c$.

3. Võtmega $(e_1, \dots, e_t)$ seotud dekrüptimisvõti $d := (e_1^{-1}, \dots, e_t^{-1})$.

Krüptosüsteem 2.2.15 (Vigenère'i šiffer) *Olgu* $\mathcal{A} = \mathbf{Z}_q$. *Vigenère'i šiffer on polüalfabeetilise substitutsioonišifri erijuht, mille korral* $e_i(p) = p + e_i \pmod q$ *(nihkešiffer!).*

Näide 2.2.16 *Olgu* $x =$ `KESEIOSANDLUGEDASEESAI` *ning* $K =$ `KASSILT`. *komponenthaaval liites saame*

$$\texttt{KESEIOSANDLUGEDASEESAI}+$$
$$\texttt{KASSILTKASSILTKASSILTK}=$$
$$\texttt{ZEETQWIKNŽÜYRÕNAETMÜTS}$$

Harjutus 2.2.17 *Olgu* $y =$ `KESEIOSANDLUGEDASEESAI` *ning* $K =$ `KASSILT`. *Leida vastav avatekst.*

Polüalfabeetilistel substitutsioonišifritel on see eelis, et krüptimisel sümbolite sagedus ei säilu. Eeltoodud näites krüptub sümbol `E` sümboliteks `A`, `S`, `T` ning `I`. Samas ei ole nende šifrite krüptoanalüüsimine oluliselt raskem. Kõigepealt tuleb leida bloki pikkus t ning seejärel leida tähtede sagedused jadades $c_1 c_{t+1} \ldots$, $c_2 c_{t+2} \ldots$ jne.

2.2.2 Permutatsioonišifrid

Krüptosüsteem 2.2.18 (permutatsioonišiffer) *Vaatleme blokkšifrit, kus bloki pikkus on* t. *Olgu* $\mathcal{P} := \mathcal{C} := \mathcal{A}^t$ *ning* $\mathcal{K} := \mathcal{S}_t$. *Suvalise võtme (permutatsiooni)* $e \in \mathcal{K}$ *jaoks defineerime*

$$E_e(p_1, \ldots, p_t) := (p_{e(1)}, \ldots, p_{e(m)}).$$

Võtmele e *seatakse vastavusse dekrüptimisvõti* $d := e^{-1}$.

Permutatsioonišiffer säilitab iga bloki siseselt erinevate sümbolite sageduse ning on seega lihtsalt krüptoanalüüsitav.

2.3 Kordšifrid

Definitsioon 2.3.1 *Olgu*

$$X_1 := (\mathcal{P}_1, \mathcal{C}_1, \mathcal{K}_1, \mathcal{E}_1, \mathcal{D}_1)$$

ning

$$X_2 := (\mathcal{P}_2, \mathcal{C}_2, \mathcal{K}_2, \mathcal{E}_2, \mathcal{D}_2)$$

kaks sümmeetrilise võtmega krüptosüsteemi, ning $\mathcal{C}_1 = \mathcal{P}_2$. *Siis öeldakse, et*

$$X_2 X_1 = (\mathcal{P}_1, \mathcal{C}_2, (\mathcal{K}_1, \mathcal{K}_2), \mathcal{E}_1 \times \mathcal{E}_2, \mathcal{D}_1 \times \mathcal{D}_2)$$

on ***kordšiffer*** *(*product cipher*), kus* $E^{(2)}_{e_2} \cdot E^{(1)}_{e_1} \in \mathcal{E}_1 \times \mathcal{E}_2$ *on kahe funktsiooni kompositsioon.*

Definitsioon 2.3.2 ***Involutsiooniks*** *nimetatakse bijektsioone, mis on iseenda pöördelemendid. Muuhulgas, kui E_e on involutsioon, siis suvalise $p \in \mathcal{P}$ korral $E_e(E_e(p) = p$.*

Harjutus 2.3.3 *Kui $E_{k_1}, \ldots, E_{k_n}$ on involutsioonid, siis $E_{k_1} \cdots E_{k_n}$ ei ole üldjuhul involutsioon. Näidata, et kehtib $(E_{k_1} \cdots E_{k_t})^{-1} = E_{k_t} \cdots E_{k_1}$.*

Substitutsiooni- ja permutatsioonišifrid eraldi ei taga eriti kõrget turvataset. Samas, nende teisenduste kombineerimisel on võimalik saavutada küllaltki suurt turvalisust. Edaspidises näeme, et mõned kõige praktilistematest ja efektiivsematest sümmeetrilisest krüptosüsteemidest on kordšifrid.

Näide 2.3.4 *Vaatleme juhtu $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbf{Z}_{32}$, olgu X_1 nihkešiffer ning X_2 — permutatsioonišiffer blokisuurusega 4, st $E_e^{(1)}(p_1p_2p_3p_4) = (p_1 + e, p_2 + e, p_3 + e, p_4 + e)$ ning $E_e^{(2)}(p_1p_2p_3p_4) = (p_{e(1)}, p_{e(2)}, p_{e(3)}, p_{e(4)})$. Seega $E^{(2)} \cdot E^{(1)} = (ep_{e_2(1)} + e_1, p_{e_2(2)} + e_1, p_{e_2(3)} + e_1, p_{e_2(4)} + e_1)$.*

Definitsioon 2.3.5 *Permutatsiooni- ja substitutsioonišifri korrutist nimetatakse* ***raundiks****. Öeldakse, et substitutsioonišiffer lisab raundile* ***konfusiooni*** *ning permutatsioonišiffer lisab* ***difusiooni****. Konfusioon teeb seose avateksti ja krüptogrammi vahel nii keerukaks kui võimalik. Difusioon tähendab avateksti bittide sellist ümber jaotamist, et suvaline liiasus avatekstiks laotuks üle terve krüptogrammi.*

Tänapäevased blokkšifrid on kordšifrid, mille igas raund peab tagama (muude eesmärkide kõrval) tingimata ka difusiooni ning konfusiooni (kuigi raunde ei saa alati lihtsalt substitutsioonideks ja permutatsioonideks lahutada).

2.4 Jadašifrid. Baasdefinitsioonid

Blokkšifrites krüptitakse järgnevad avateksti blokid, kasutades iga kord sama võtit e. See tähendab, et krüptogramm c saadakse järgnevalt:

$$c := c_1c_2 \ldots = E_e(p_1)E_e(p_2) \ldots .$$

Alternatiivne võimalus oleks kasutada nn. **jadašifreid** (*stream cipher*), kus genereeritakse **võtmejada** (*keystream*) $z := z_1z_2 \ldots$ ning krüptitakse avatekst järgmiselt:

$$c := c_1c_2 \ldots = E_{z_1}(p_1)E_{z_2}(p_2) \ldots .$$

Olgu $e \in \mathcal{K}$ ning $p := p_1p_2 \ldots \in \mathcal{P}^*$. Võtmejada element

$$z_i := f_i(e, p_1, \ldots, p_{i-1}),$$

st. et z_i on funktsioon võtmest e ning avateksti p esimesest $i - 1$ sümbolist. Seega jadašifrites arvutatakse üksteise järel z_1, c_1, z_2, c_2, ... Dekrüptimisel arvutatakse järjest z_1, p_1, z_2, p_2, ... Ehk formaalselt

Definitsioon 2.4.1 ***Jadašifriks*** *nimetatakse seitsmikut*

$$(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{L}, \mathcal{F}, \mathcal{E}, \mathcal{D}),$$

kus

1. *$\mathcal{P}$ on võimalike avatekstide lõplik hulk.*
2. *$\mathcal{C}$ on võimalike krüptogrammide lõplik hulk.*
3. *$\mathcal{K}$ on võimalike võtmete lõplik hulk.*
4. *$\mathcal{L}$ on lõplik* ***võtmejada tähestik*** *(*keystream alphabet*).*
5. *$\mathcal{F} = (f_1, f_2, \dots)$ on* ***võtmejada generaator*** *(*keystream generator*). Iga $i \geq 1$ jaoks $f_i : \mathcal{K} \times \mathcal{P}^{i-1} \to \mathcal{L}$.*
6. *Suvalise $z \in \mathcal{L}$ jaoks leiduvad reeglid $E_z \in \mathcal{E}$ ning $D_z \in \mathcal{D}$, $E_z : \mathcal{P} \to \mathcal{C}$ ning $E_z : \mathcal{C} \to \mathcal{P}$, nii et $D_z(E_z(x)) = x$ suvalise $x \in \mathcal{P}$ korral.*

Jadašifrit nimetatakse **sünkroonseks**, kui võtmejada ei sõltu avatekstist. Sellisel juhul võib võtit e vaadelda kui "seemet" (*seed*), mida laiendatakse võtmejadaks $z_1 z_2 \dots$. Jadašifrit nimetatakse **perioodiliseks** perioodiga t, kui $z_{i+t} = z_i$ iga täisarvu $i \geq 1$ korral.

Näide 2.4.2 *Ka Vigènere'i šiffer on perioodiline jadašiffer perioodiga t, võtmeks on $e = (e_1, \dots, e_t)$ ning $z_i = e_i$ iga i korral, kusjuures $E_z(p) := p + z$ (võrdle nihkešifriga).*

Lühikese perioodiga võtmejadad (näiteks Vigènere'i šiffer) on kergesti lahti murtavad.

Jadašifreid vaadeldakse sageli üle tähestiku $\mathcal{P} = \mathcal{C} = \mathcal{L} = \mathbf{Z}_2$, kus $E_z(p) := p \oplus z$ ning $D_z(c) := c \oplus z$ (praktika: välistav või on riistvaras efektiivselt realiseeritav).

Sünkroonse võtmejada genereerimist saab kirjeldada ka järgnevalt. Olgu meil olemas jada $(e_1, \dots, e_t)$ ning $z_i := e_i$, $1 \leq i \leq t$. Defineerime

$$z_{i+t} := \sum_{j=0}^{t-1} c_j z_{i+j} \mod 2,$$

kus $c_i \in \mathbf{Z}_2$ on mingid konstandid. Sellist võrdust nimetatakse *t-astme lineaarseks rekurrentsusseoseks*: t-astme, kuna iga liidetav sõltub eelmisest t-ist liidetavast; lineaarne, kuna z_{i+t} on eelmiste termide lineaarne funktsioon. Me võime eeldada, et $c_0 = 1$, kuna vastasel korral oleks tegu $(t-1)$-astme võrrandiga. Väikseimat t-i, mille korral leidub seda võtmejada genereeriv t-astme lineaarne rekurrentsusseos, nimetatakse antud võtmejada **lineaarseks keerukuseks**.

Harjutus 2.4.3 *Siin on võtmeks 2t väärtust* $e_1, \dots, e_t, c_1, \dots, c_{t-1}$. *Kui*

$$(e_1, \dots, e_t) = (0, \dots, 0),$$

siis on terve võtmejada nulline ja sel juhul on krüptogramm identne avatekstiga. Näidata, et iga teine algvektor $(e_1, \dots, e_t)$ *genereerib (sobivate konstantide* c_i *korral) perioodilise võtmejada perioodiga* $2^t - 1$*: st. lühikesest võtmest saame päris pika võtmejada perioodi.*

Võtmejada sellisel genereerimisel on veel üks positiivne aspekt: nimelt on säärast genereerimist riistvaraliselt väga lihtne teostada, kasutades **lineaarse tagasisidega nihkeregistrit** (*linear feedback shift register*, LFSR). LFSR initsialiseeritakse vektroriga $(\ell_1, \dots, \ell_t) := (k_1, \dots, k_t)$, igal sammul tehakse järgmist:

1. väljastatakse ℓ_1 kui võtmejada järgmine bitt.
2. $i > 1$ korral nihutatakse ℓ_i üks bitt vasakule.
3. arvutatakse $\ell_t := \sum_{j=0}^{t-1} c_j \ell_{j+1}$.

Näide 2.4.4 *Olgu* $t = 4$ *ning võtmejada genereerimiseks kasutatakse reeglit* $z_{i+4} := z_i \oplus z_{i+1}$. *Suvalise nullvektorist erineva algvektori korral on tekkiva võtmejada perioodiks 15. Näiteks, alustades vektorist* $(1, 0, 0, 0)$, *saame võtmejada* $1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, \dots$. *Suvaline teine nullist erinev algvektor genereerib antud võtmejada tsüklilise permutatsiooni. Vastav LFSR on kujutatud joonisel 2.2.*

lfsr.eps

Joonis 2.2: Lineaarse tagasisidega nihkeregister.

Mittesünkroonse nihkešifri näitena vaatleme järgmist šifrit.

Krüptosüsteem 2.4.5 (*autokey cipher*) *Olgu* $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathcal{L} = \mathbf{Z}_{32}$. *Olgu* $z_1 := e$ *ning* $z_i := x_{i-1}$, $i \geq 2$. *Suvalise* $0 \leq z \leq 31$ *korral*

$$e_z(p) := p + z \mod 32 \text{ ja}$$
$$d_z(c) := c - z \mod 32,$$

kus $p, c \in \mathbf{Z}_{32}$.

Suvaline madala lineaarse keerukusega jadašiffer on lihtsalt lahti murtav ning seega tuleks vältida nende kasutamist reaalsetes rakendustes. Jadašifrite uurimise üheks motivatsiooniks on järgmine šiffer.

Krüptosüsteem 2.4.6 (Kordumatu numbrijada (*one-time pad*)) .
Olgu

$$\mathcal{P} = \mathcal{K} = \mathcal{C} = \mathbf{Z}^t,$$

$t > 0$. *Defineerime*

$$E_e(p) := p \oplus e$$
$$D_e(c) := c \oplus e.$$

Kordumatu numbrijada on tingimusteta turvaline:

$$(\forall c \in \mathcal{C})(\forall p \in \mathcal{P})(\exists e \in \mathcal{K}) c = p \oplus e.$$

Halvaks omaduseks on see, et võti on sama pikk kui avatekst ning *iga* uue avateksti šifreerimiseks tuleks turvalisuse saamiseks valida uus võti.

Jadašifrid pakuvad siin ühe lahenduse, lühikesest võtmest genereeritakse pikk (pseudojuhuslik) võtmejada, mis lõppkokkuvõttes xor'itakse avatekstiga läbi. Sellise võtmejada genereerimiseks läheb meil vaja krüptograafiliselt tugevat **pseudojuhuslike arvude generaatorit** (*pseudo-random number generator*).

Ajalooliseks ülevaateks lihtšifritest võib soovitada eesti keeles ilmunud raamatut [Liv94]. Põhjalikku ülevaadet jadašifrite teooriast võib leida artiklist [Rue92].

2.5 Krüptoanalüüs

Tuletagem meelde, et krüptoloogia koosneb krüptograafiast ja krüptoanalüüsist ning krüptoanalüütikud püüavad kõigest hingest lahti murda krüptograafide poolt koostatud krüptosüsteeme (muuseas enamus krüptograafe on samal ajal ka krüptoanalüütikud).

Krüptoanalüüsi raskus sõltub oluliselt sellest, kas analüütik saab ka aktiivselt rünnata, või on ta ainult passiivne kuulaja. Krüptosüsteemi ründeid jaotatakse (tavaliselt) nelja alaliiki, kus iga järgmine rünne on eelmisest ohtlikm):

1. **krüptogrammirünne** (*ciphertext-only*), kus Oskar teab mingit krüptogrammi y.

2. **teadaoleva avatekstiga rünne** (*known plaintext*), kus Oskar teab mingit avateksti ja sellele vastavat krüptogrammi.

3. **valitava avatekstiga rünne** (*chosen plaintext*), kus Oskaril on ajutine ligipääs krüptimisprotsessile: Oskar võib vabalt valida avateksti x ning konstrueerida sellele vastava krüptogrammi.

täht	sagedus	täht	sagedus
A	0.082	N	0.067
B	0.015	O	0.075
C	0.028	P	0.019
D	0.043	Q	0.001
E	0.127	R	0.060
F	0.022	S	0.063
G	0.020	T	0.091
H	0.061	U	0.028
I	0.070	V	0.010
J	0.002	W	0.023
K	0.008	X	0.001
L	0.040	Y	0.020
M	0.024	Z	0.001

Joonis 2.3: Inglise keele tähtede sagedustabel

4. **valitava krüptogrammiga rünne** (*chosen ciphertext*), kus Oskaril on ajutine ligipääs dekrüptimisele: ta saab valida krüptogrammi y ning konstrueerida sellele vastava avateksti.

Järgnevas näites eeldame, et meil on tegu ingliskeelse tekstiga, kus ei kasutata tühikuid ja kirjavahemärke, nii et tüüpiline teade võiks välja näha järgmine: IWISHIWASALITTLEBITTALLERIWISHIWASABALLER. Tänu sellele toimub tegevus ka ringis $\mathbf{Z}_{26}$, kus tähed ja numbrid on vastavusse seatud järgmiselt:

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

Vaatleme nõrgimat, krüptogramm-rünnet, monoalfabeetiliste krüptosüsteemide korral ja näitame, et isegi see rünne on edukas.

Meie lähenemine baseerub tähtede sagedusanalüüsil, kasutades inglise keele liiasusi, nimelt seda, et standardses ilukirjanduslikus tekstis

1. Iga täht esineb teatud sagedusega (vt joonis 2.3), seejuures paneme tähele, et tähed jaotuvad sageduselt viide rühma: 1. E (0.120) 2. T, A, O, I, N, S, H, R (0.06 ja 0.09 vahel) 3. D, L (umbes 0.04) 4. C, U, M, W, F, G, Y, P, B (0.015 ja 0.023 vahel) 5. V, K, J, X, Q, Z (vähem kui 0.01).

2. Iga tähepaar ehk **digramm** esineb teatud sagedusega; 30 kõige sagedasemat digrammi inglise keeles (sageduse langemise järjekorras) on: TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU, EA, NG, AS, OR, TI, IS, ET, IT, AR, TE, SE, HI, OF.

täht	sagedus	täht	sagedus
A	2	N	1
B	1	O	1
C	0	P	3
D	6	Q	0
E	5	R	8
F	4	S	3
G	0	T	0
H	5	U	2
I	0	V	4
J	0	W	0
K	5	X	2
L	2	Y	1
M	2	Z	0

Joonis 2.4: Tähtede esinemiste arv

3. Sama kehtib ka tähekolmikute ehk **trigrammide** kohta, analoogiline nimekiri oleks: `THE`, `ING`, `AND`, `HER`, `ERE`, `ENT`, `THA`, `NTH`, `WAS`, `ETH`, `FOR`, `DTH`.

Olgu käsiteldavaks süsteemiks afiinne šiffer ning krüptogrammiks ([Sti95], paragrahv 1.2.1):

`FMXVEDKAPHFERBNDKRXRSREFMORUDSDKDVSHVUFEDKAPRKDLYEVLRHHRH`

Selle krüptogrammi tähtede esinemiste sagedus on toodud tabelis 2.4. Meie krüptogrammis on ainult 57 tähte, kuid sellest täiesti piisab, et võtit leida.

Kõige sagedasemad tähed on `R` (8 esinemist), `D` (6 esinemist), `E`, `H` ja `K` (igal 5 esinemist) ning `F`, `S` ja `V` (igal 4 esinemist). Kuna kõige sagedam täht on `R`, oleks esimeseks hüpoteesiks[1], et `R` on `e` ja `D` on `t` šifreering, kuna `e` ja `t` on kaks kõige tavalisemat tähte. Arvuliselt väljendades, $E_e(4) = 17$ ja $E_e(19) = 3$. Saame kaks lineaarvõrrandit kahe tundmatuga:

$$4\alpha + \beta = 17$$
$$19\alpha + \beta = 3,$$

sellel süsteemil on ringis $\mathbf{Z}_{26}$ unikaalne lahend $\alpha = 6$, $\beta = 19$. Kuid see lahend ei sobi, sest $\gcd(\alpha, 26) = 2 > 1$. Seega on esimene hüpotees väär.

Oletame nüüd, et `R` on `e` ja `E` on `t` šifreering. Toimides samamoodi kui eespool, saame $\alpha = 13$, ent seegi vastus ei sobi. Proovime kolmandat võimalust, et `R` on `e` ja `H` on `t` šifreering. Siit saame $\alpha = 8$, jällegi võimatu. Olgu siis `R` — `e` ja `K` — `t` šifreering. Siit saame $\alpha = 3$ ja $\beta = 5$ ning see on vähemalt lubatud võti ($\gcd(\alpha, 26) = 1$). Proovime antud krüptogrammi seda võtit kasutades dešifreerida ning vaatame, kas saame tulemuseks "normaalse" ingliskeelse teksti;

[1] siinkohal kirjutame avateksti sümboleid väikeste ja krüptogrammi omi suurte tähtedega

kui saame, siis on ilmselt võti ka õige (ka inglise keel on küllalt suure liiasusega, st enamus tähestringe ei tähenda mitte midagi). Kuna $(\alpha, \beta) = (3, 5)$, siis $D_K(y) = 9y - 19$ ning antud krüptogrammi dešifreerimine annab meile:

```
algorithmsarequitegeneraldefinitionsofarithmeticprocesses
```

Seega, nagu nägime, pika krüptogrammi puhul ei ole afiinse šifri lahtimurdmine eriti suur probleem. Sama kehtib kõikide siiamaani vaadeldud šifrite korral.

Tegelikult on teada, et sümmeetriliste krüptosüsteemide kasutamisel leidub alati selline (krüptosüsteemist sõltuv) $m = m(k)$ (kus k on võtmepikkus), nii et Oskar saab kõik pikkusega $|y| > m$ krüptogrammid dešifreerida üheselt ([Sti95], teoreem 2.11). Püüame seda väidet seletada.

2.5.1 Informatsiooniteooria

Niinimetatud **Shannoni mudelis** on krüptoanalüütiku (Oskari) käsutuses limiteerimata aeg ja arvutusvõimsused. Fikseerime krüptosüsteemi ning avateksti ja võtme tõenäosusjaotused $p(\mathcal{P})$ ja $p(\mathcal{K})$ vastavalt hulkades $\mathcal{P}$ ja $\mathcal{K}$. Shannoni mudelis on krüptoanalüüs defineeritud kui avateksti (või võtme) leidmine, kui teada on vaid krüptogramm ning avatekstide ning võtmete jaotus. Mudelit loetakse lahti muugituks, kui leitakse "ühene" krüptogramm, st selline, mille tõenäosus on praktiliselt üks ning samas kõikide teiste krüptogrammide tõenäosused on praktiliselt nullid.

Vaatleme tõenäosusi

1. $P(Y)$: antud jaotuste korral $E_K(X) = Y$.
2. $P(Y|K)$: suvalise $X \in \mathcal{P}$ ja antud K korral $E_K(X) = Y$.

Sammuke abstraktsuse poole.

Definitsioon 2.5.1 *Olgu $\boldsymbol{X}$ juhuslik suurus üle lõpliku hulga tõenäosusjaotusega $p(\boldsymbol{X})$. Selle jaotuse **entroopiaks** nimetatakse suurust*

$$H(\boldsymbol{X}) := -\sum_{i=1}^{n} p_i \log_2 p_i,$$

kus i muutub üle kõigi indeksite i, $p_i > 0$.

Informaalselt on $H(\mathbf{X})$ meie "ebakindlus" tulemis. Analoogselt tõenäosustega defineeritakse ka suurus $H(\mathbf{X}|\mathbf{Y})$. Tähistame $I(\mathbf{X}, \mathbf{Y}) := H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y})$. Konkreetseks tagasi minnes olgu X^n ja Y^n n-bitised avatekst ja krüptogramm.

1. Kui $I(X^n, Y^n) = 0$ suvalise n korral (st X^n ja Y^n on statistiliselt sõltumatud), siis öeldakse, et antud krüptosüsteem on **perfektselt turvaline**. Krüptoanalüütik saab avateksti leidmise juures kasutada vaid erinevate avatekstide esinemistõenäosusi. Perfektsetes krüptosüsteemides $H(K) \geq H(X^n)$, st et uut võtit on vaja iga konstantse ajavahemiku järel.

Perfektse krüptosüsteemi avateksti lõigu teadmine võib meile kätte anda antud avateksti krüptimisel kasutatud võtme lõigu, kuid ei muuda üldjuhul ülejäänud avatekstilõikude esinemistõenäosusi. Kordumatu numbrijada on perfektne krüptosüsteem.

2. Suurte n-ide korral $0 < I(X^n, Y^n) < H(X^n)$. Sel juhul jääb alles teatud ebakindlus avateksti suhtes, mida ei õnnestu kõrvaldada ka limiteerimata arvutusvõimsuse korral. Lõpliku võtmega krüptosüsteeme, mille korral $H(K|Y^n) > 0$ kui $n \to \infty$, nimetatakse **ideaalselt turvalisteks**. Krüptoanalüütiku jaoks jaotused täpsustuvad, kuid X^n-l ei ole olemas ühest lahendust. See ei pruugi tähendada, et krüptosüsteem oleks turvaline, krüptoanalüütik võib jääda ebakindlaks vaid avateksti mingi piiratud lõigu suhtes. Ideaalse turvalisuse saavutamine nõuab tavaliselt perfektset andmete pakkimist (**aritmeetiline kodeerimine**), samuti sõltub see kriitiliselt avateksti täielikust salastusest.

3. Suurte n-ide korral $I(X^n, Y^n) = H(X^n)$. Pärast teatavat arvu krüptogrammi bitte (tüüpiliselt piisab $H(K)/(1-h)$ bitist, kus h on avateksti bittide keskmine liiasus) jääb avateksti ja tollele rakendatud võtme jaoks alles unikaalne võimalus. Nö praktiliste krüptosüsteemide juhtum.