CYBERNETICA

Avalik

# Privaatsuskaitse tehnoloogiate kontseptsioon

Aruanne

Versioon 1.1

31.03.2023

ID D-16-175

Projektijuhid: Liina Kamm (Cybernetica AS)
Nele Nisu (Majandus- ja Kommunikatsiooniministeerium)

Autorid: Dan Bogdanov (Cybernetica AS)
Eduardo Brito (Cybernetica AS)
Paula Etti (Cybernetica AS)
Liina Kamm (Cybernetica AS)
Peeter Laud (Cybernetica AS)
Tanel Mällo (Cybernetica AS)
Andre Ostrak (Cybernetica AS)
Kati Sein (Cybernetica AS)
Riivo Talviste (Cybernetica AS)
Maria Toomsalu (Cybernetica AS)

Cybernetica AS, Mäealuse 2/1, 12618 Tallinn, Estonia.

E-post: info@cyber.ee, Veebileht: https://www.cyber.ee, Telefon: +372 639 7991.

# Sisukord

# 1 Sissejuhatus

## 1.1 Motivatsioon

Mis oleks, kui Eesti oskaks ehitada mitmete inimeste või organisatsioonide andmete peal põhinevaid teenuseid sama turvaliselt kui üksikisiku teenuseid üle X-tee? Kui andmete töötlemisel oleks tagatud läbipaistvus ja privaatsus? Kui andmeid oleks võimalik erinevate osapoolte vahel töödelda ilma üksikisikule viitavaid andmeid avaldamata? Millised uued võimalused sellisest võimekusest Eestile avaneda võiksid?

Eesti on tuntud kui riik, kus rakendatakse laialdaselt tehnoloogiat. Andmete kogumise ja töötlemise suurenemisega muutuvad privaatsusküsimused üha olulisemaks.

Turvaline andmekasutus, näiteks privaatsuskaitse tehnoloogiate rakendamine lõimitud andmekaitse osana, võimaldaks arendada uusi kvaliteetseid e-riigi teenuseid ja viiks ühtlasi lähemale inimkesksemale e-valitsemisele. Nii on võimalik muuta protsesse efektiivsemaks ja hoida kokku nii ajalist kui ka rahalist ressurssi. Näiteks oleks võimalik määrata inimestele toetuseid vajaduspõhiselt vastavalt nende tegelikele kuludele, ühendades selleks avaliku ja erasektori andmeid.

Lisaks eelnevale võimaldaks turvaline andmekasutus ja sellel põhinev andmete analüüs tuge riigijuhtidele läbimõeldud otsuste langetamiseks.

Veelgi enam, teaduskoostöö uute tervishoiu-, finants- või muude teenuste loomiseks riigi, teadusasutuste ja ettevõtete vahel edeneks nii Eesti sees, Euroopas kui ka väljaspool. Erasektoris ja teadusasutustes oleks tänasega võrreldes võimalik oluliselt ulatuslikumalt töödelda andmeid, tagades inimeste privaatsus, uuringuteks, teenuste osutamiseks ja andmepõhiseks otsustamiseks.

See pakuks ainest ka uute äriideede sünniks ja aitaks kaasa uute ettevõtete loomisele, aidates kaasa uute e-residentide ja ettevõtete tekkimisele, tooks riigile maksutulu ja välisinvesteeringuid ning hoiaks Eesti tugevat digiriigi kuvandit.

Lisaks toetaks privaatsusküsimuste lahendamine andmemajanduse kasvu läbi ulatuslikuma andmete kasutuse ja väärindamise. Üldiselt võib privaatsuskaitse tehnoloogiate rakendamine Eestis viia turvalisema ja usaldusväärsema digitaalse keskkonna loomiseni, mis soodustab innovatsiooni, digiriigi arengut ja ettevõtlust, samal ajal kaitstes inimeste õigusi.

Nende unistuste täitmiseks on teadlased ja insenerid aastaid arendanud privaatsuskaitse tehnoloogiaid (ingl k **privacy enhancing technologies**, edaspidi lühendina PET).

## 1.2 Uuringu lühitutvustus

2022. aastal algatatud Eesti privaatsuskaitse tehnoloogiate uuringul on kaks väljundit.

1. **Privaatsuskaitse tehnoloogiate kontseptsioon** (see dokument) kirjeldab tehnoloogiaid ja pakub nende rakendamiseks e-riigis üldised mudelid ning kontseptsiooni.
2. **Privaatsuskaitse tehnoloogiate Eestis rakendamise teekaart** kirjeldab intervjuude põhjal Eesti avaliku sektori asutuste kogemusi privaatsuskaitse tehnoloogiatega, nende vajadusi ning pakub välja arenduse teekaardi 2023. aasta seisuga.

Uuringu lugeja võib valida, millisest dokumendist ta uuringu tulemitega tutvumist alustab. Rakendamise kontseptsioonist alustaja saab teada, kuidas privaatsuskaitse tehnoloogiad töötavad,

kuidas neid teistes riikides kasutatud on ning milliseid e-riigi probleeme need lahendavad. Mis veelgi tähtsam – rakendamise kontseptsioon kirjeldab, kuidas organisatsioon saab oma arendustsüklis privaatsuskaitse tehnoloogiaid kasutusele võtta.

Teekaardist alustades saab lugeja kõigepealt teada, milliseid privaatsuskaitse tehnoloogiaid Eestis kasutatud on ja millised on avaliku sektori organisatsioonide vajadused. Dokumendi lõpus olev arendusplaan aitab koostada tööplaane ning teha otsuseid investeeringute vajaduse kohta.

## 1.3 Dokumendi käsitlusala

Privaatsuskaitse tehnoloogiad[1] (ingl k *privacy enhancing technologies*, PET) on info- ja sidetehnoloogilised meetmed, tooted või teenused, mis kaitsevad privaatsust andmete välistuse või vähendamisega või isikustatavate andmete tarbetu ja/või soovimatu töötluse vältimisega, samas säilitades süsteemi võimed. **PETide rakendamine võib kaitsta ka riikide, ettevõtete ja kõikvõimalike teiste osapoolte andmeid, sh konfidentsiaalset informatsiooni ja äri- või riigisaladusi.**

Laiemas plaanis on oluline silmas pidada, et vahel peetakse privaatsust ja andmekaitset (või infoturvet) sünonüümideks ning seetõttu nähakse andmete kogumise ja/või töötlemise kontrollimist, välistamist või vähendamist peamiste privaatsuskaitse lahendustena. Samas on oluline pöörata tähelepanu eetika ja isiku autonoomia aspektidele, mida infoturbenõuded ei kata.

Euroopa Liidu põhiõiguste harta[2] eristab privaatsust ja andmekaitset selgelt (vastavalt harta artiklid 7 ja 8). Lühidalt lähtub privaatsus reeglina üksikisiku vaatenurgast, kelle huvi on vältida liigset sekkumist tema eraellu. Andmekaitse seostub samas pigem organisatsioonide huviga käitada enda toimimiseks vajalikke protsesse kooskõlas kehtivate seadustega, et minimeerida regulatsioonidega vastuollu minekuga kaasnevaid äririske.

Terviklikumalt hõlmab erinevate osapoolte huve privaatsuskaitse tehnoloogiate aluseks olev privaatsuseesmärkide käsitlus, mis lisaks peamiselt andmekaitsega seonduvale andmete seostamatusele (isikustatavuse välistamine/vähendamine, ingl k *unlinkability*) hõlmab ka andmetöötluse läbipaistvuse (ingl k *transparency*) ja sekkutavuse (inimese kui andmesubjekti võimalused andmetöötluse katkestamiseks, ingl k *intervenability*) tagamist [1].

Neid kolme privaatsuseesmärki tuleb käsitleda eraldiseisvalt kolmest klassikalisest teabe turvalisuse komponendist, milleks on konfidentsiaalsus (ingl k *confidentiality*), terviklus (ingl k *integrity*) ja käideldavus (ingl k *availability*). Nii tagatakse, et konkreetne IT-süsteem arvestab paremini oma organisatsiooniliste ja sotsiaalsete dimensioonidega, st ümbritseva maailmaga ning inimkesksusega.

Privaatsuskaitse tehnoloogiate arendamisel on eesmärgina kõige suuremat tähelepanu pälvinud andmete seostamatus, kuna see toimib andmete väärkasutamise esmase kaitseliinina [1] ning hõlmab ühtlasi rakenduslikus mõttes kõige küpsemaid tehnoloogiaid.

Standardtingimustel korrektselt juurutatuna vähendavad andmete seostamatust tagavad lahendused olulisel määral privaatsusriive riske [2]. Seetõttu käsitleb ka käesolev ülevaade esmajärjekorras just neid privaatsuskaitse tehnoloogiaid. Läbipaistvust ja sekkutavust võimaldavate lahenduste ehitamine tähendab loodavale teenusele uute kasutuslugude ja funktsionaalsuse li-

[1] Erinevates eestikeelsetes dokumentides kasutatakse ka termineid privaatsust säilitavad tehnoloogiad ning privaatsustehnoloogiad. Selguse huvides kasutame selles aruandes läbivalt terminit privaatsuskaitse tehnoloogia(d), mis on käsitletav eelnimetatud terminite sünonüümina, kuid on mõistena tunnustatud ka andmekaitse ja infoturbe leksikonis AKIT. – Internetis: https://akit.cyber.ee/term/13759-privaatsuskaitse-tehnoloogia (viimati külastatud 01.03.2023).

[2] Euroopa Liidu põhiõiguste harta, https://fra.europa.eu/et/eu-charter (viimati külastatud 01.03.2023).

samist. Nende funktsioonide ehitamiseks on olemas toetavaid tehnoloogiaid, kuid ennekõike peavad nad olema tugevalt integreeritud loodava süsteemi või teenuse talitlusloogikaga.

**Eelnevast lähtudes on selles aruandes fookus privaatsuskaitse tehnoloogiatel, mis tagavad seostamatust.** Läbipaistvuse ja sekkutavuse toetamiseks vajalikke tehnoloogiaid on loetletud vähemal määral. Nende kahe privaatsuseesmärgi saavutamiseks soovitame läheneda läbi süsteemide ja teenuste arendusprotsessi.

## 1.4 Mõisted

**konfidentsiaalsus (ingl k *confidentiality*)**
: üks teabe turvalisuse kolmest põhikomponendist; andmete omadus, mis näitab, mil määral need andmed ei ole volitamata isikutele, protsessidele või muudele olemitele kättesaadavad ega avalikustatud;

**käideldavus (ingl k *availability*)**
: üks teabe turvalisuse kolmest põhikomponendist; omadus olla volitatud olemi nõudel õigel ajal kättesaadav ja kasutuskõlblik;

**lõimitud andmekaitse (ingl k *data protection by design*)**
: asjakohaste tehniliste ja korralduslike meetmete rakendamine nii, et saab tõhusalt rakendada andmekaitsepõhimõtteid; vajalike kaitsemeetmete lõimimine isikuandmete töötlemisse, et täita Euroopa Liidu isikuandmete kaitse üldmääruse (edaspidi IKÜM)[3] nõudeid ja kaitsta andmesubjektide õiguseid;

**läbipaistvus (ingl k *transparency*)**
: süsteemi või protsessi avatus ja jälitatavus; omadus, mis tagab, et kogu privaatsust puudutav andmetöötlus, sealhulgas õiguslik, tehniline ja korralduslik külg oleks arusaadav ja ennistatav;

**privaatsuskaitse tehnoloogia(d) (ingl k *privacy enhancing technologies*)**
: info- ja sidetehnoloogilised meetmed, tooted või teenused, mis kaitsevad privaatsust isikustatavate andmete välistuse või vähendamisega või isikustatavate andmete tarbetu ja/või soovimatu töötluse vältimisega, samas säilitades süsteemi võimed;

**privaatsuslõime (ingl k *privacy by design*)**
: süsteemitehniline käsitlusviis, mis arvestab privaatsust kogu süsteemi elutsüklis ning näeb ette privaatsusnõuete esitamist süsteemide, tehnoloogiate, äritavade jms spetsifikatsioonides, võrdle ka mõistega ”lõimitud andmekaitse”;

**privaatsustehnika (ingl k *privacy engineering*)**
: distsipliin, mis keskendub juhistele, kuidas vähendada privaatsusriske ning põhjendada otsuseid ressursside paigutamiseks ja meetmete toimivaks teostuseks infosüsteemides;

**sekkutavus (ingl k *intervenability*)**
: omadus, mis tagab, et privaatsust puudutavasse kogu andmetöötlusse saavad sekkuda kohaldatavate õigusaktide tingimustel andmesubjektid, isikutuvastusteabe korraldajad, isikutuvastusteabe töötlejad, järelevalveorganid;

**seostamatus (ingl k *unlinkability*)**
: ründe või privaatsuse kontekstis on huviobjektid süsteemis seostamatud, kui nende uurimine väljastpoolt ei anna lisateavet nende võimaliku seotuse kohta;

[3] Euroopa Parlamendi ja nõukogu määrus (EL) 2016/679, 27. aprill 2016, füüsiliste isikute kaitse kohta isikuandmete töötlemisel ja selliste andmete vaba liikumise ning direktiivi 95/46/EÜ kehtetuks tunnistamise kohta (isikuandmete kaitse üldmäärus) [3].

**terviklus (ingl k *integrity*)**
üks teabe turvalisuse kolmest põhikomponendist; varade õigsuse ja täielikkuse kaitstus;

**vaikimisi andmekaitse (ingl k *data protection by default*)**
tehnilised ja korralduslikud meetmed tagamaks, et töödeldaks üksnes iga konkreetse töötlusotstarbe jaoks vajalikke isikuandmeid.

## 1.5 Alusmaterjalid

Aruande koostamisel lähtusime järgmistest materjalidest:

1. Arenguseire Keskus, *Andmeühiskonna tulevik. Stsenaariumid aastani 2035* [4];
2. Majandus- ja Kommunikatsiooniministeerium, *Eesti digiühiskond 2030* [5];
3. Ühendkuningriigi akadeemia *The Royal Society* aruanne *From privacy to partnership: The role of privacy enhancing technologies in data governance and collaborative analysis* [6];
4. Ühendkuningriigi andme-eetika ja -innovatsioonikeskuse CDEI aruanne *Privacy Enhanching Technologies Adoption Guide* [7];
5. ÜRO komitee *United Nations Committee of Experts on Big Data and Data Science for Official Statistics* aruanne *United Nations Guide on Privacy-Enhancing Technologies for Official Statistics* [8];
6. Erinevate valdkondade teaduskirjandus (täielik ülevaade dokumendi lõpus bibliograafias).
7. Erinevad standardid (täielik ülevaade dokumendi lõpus bibliograafias).

## 1.6 Aruande struktuur

Peatükk 2 kirjeldab privaatsuskaitse tehnoloogiate kasutamise vajadust ning selle õiguslikke, tehnilisi ja sotsiaalseid aspekte. Aruanne lähtub Euroopa Liidu õigusruumist ning Eesti digiühiskonna arengukavast.

Peatükk 3 selgitab, kuidas privaatsuskaitse tehnoloogiaid ning privaatsustehnikat üldiselt siduda infosüsteemide ja teenuste elutsükliga. Oleme toetunud eeldusele, et digiühiskonnas on infosüsteemid oma elutsüklite erinevatel etappidel – mõned neist on juba küpsed ning teised alles kavandamisel. Peatükis kirjeldatud lähenemisviis on sobilik neile kõigile.

Peatükis 4 kirjeldame aruandesse välja valitud privaatsuskaitse tehnoloogiaid ning nende õiguslikke aspekte. Lisaks põhjalikule tekstilisele esitusele on peatükis ka iga tehnoloogia kohta üheleheküljeline kokkuvõte (koos joonisega), mida aruande lugejad saavad kasutada oma teadmusbaasi täiendamiseks, aga ka tehnoloogiate hindamise protsessis. Valitud tehnoloogiate jaoks toome välja inspireerivad näited reaalse maailma rakendustest. Ka need ülevaated on aruandes esitatud lihtsustatud kujul, üheleheküljeliste illustreeritud tabelitena.

Peatükk 5 annab ülevaate privaatsuskaitse tehnoloogiate juurutamise kogemustest teistes maailma riikides. Ülevaates tuuakse välja arendusprogrammid, uuringud ja märkimisväärsemad rakendused. Peatüki eesmärk on jagada edulugusid ja kogemusi ning anda inspiratsiooni uute rakenduste loomiseks Eestis.

Peatükis 6 esitame teenusemudeleid, mis saavad privaatsuskaitse tehnoloogiate kasutuselevõtuga paremad andmekaitsegarantiid. Need teenusemudelid täidavad privaatsuseesmärke – seostamatust, läbipaistvust ja sekkutavust – paremini, kui aruande kirjutamise hetkel levinud tehnoloogiad.

# 2 Millist kasu saab digiühiskond privaatsustehnoloogiate rakendamisest?

## 2.1 Privaatsustehnoloogiad on digiühiskonna arengu üks alus

Eesti Vabariigi digiühiskonna arengukava aastani 2030 näeb inimeste põhiõiguste, sh privaatsuse kaitset ühe põhimõttena [5]. Digiühiskonna üheks järgmiseks arenguhüppeks loeb arengukava inimkeskse digiriigi arengut. Konkreetsete tegevustena on plaanis andmejälgija ja nõusolekuteenuse laiem juurutamine ning teadlikkuse tõstmine. Eraldi eesmärkideks on veel digilahenduste usaldusväärsuse ning inimkesksuse tagamiseks vajalike riskihaldusmeetmete väljatöötamine ja digiteenuste omanike ja ehitajate suutlikuse arendamine, et nad suudaks pakkuda inimkeskseid ja usaldusväärseid digilahendusi. Need tegevused aitavad omavahelises koostöös senisest paremini saavutada läbipaistvuse ja sekkutavuse privaatsuseesmärke.

Avaramas, ühiskondlikus vaates saab indiviid nii võimaluse oma privaatsuse tagamisel aktiivse osapoolena tegutseda. Tekkiv avatud digiühiskond on eraldi väärtus ja indiviidist võib selles saada innovatsiooni toetaja.

Lisaks peab arengukava privaatsuskaitse tehnoloogiate rakendamist oluliseks andmepõhise riigivalitsemise ja andmete taaskasutuse saavutamisel. Konkreetse tegevusena plaanitakse riikliku privaatsuskaitse tehnoloogiate rakendamise programmi elluviimist. Nende tehnoloogiate rakendamisega saavutame parema läbipaistvuse ja sekkutavuse ning senisest palju tugevamal tasemel seostamatuse.

## 2.2 Privaatsustehnoloogiad võimaldavad uute teenuste loomist

Andmete taaskasutuses nähakse suurt potentsiaali nii üksikisiku, riigiasutuste kui erasektori teenuste arenduses kui ka andmepõhiste otsuste tegemisel. Üksikisikule saab tema andmete põhjal pakkuda sündmusteenuseid, vajaduspõhiseid toetuseid või ka personaalseid teenuseid.

Avalikus sektoris loob andmete laialdasem kasutus võimaluse ressursside efektiivsemaks kasutuseks. Seda tänu menetluste automatiseerimisele ja andmetel põhinevale kvaliteetsemale poliitikaotsuste kujundamisele. Erasektorile võivad avaliku sektori käsutuses olevad isikuandmed olla usaldusväärseks allikaks, millele tuginedes ehitada uusi teenuseid.

Mõningad olulisemad valdkonnad, mis täna privaatsuskaitse tehnoloogiate süsteemsest rakendamisest võidaksid, on tervishoid, finants, haridus, internetiturundus, asjade internet ja avalik haldus. Toome ühe näite tervishoiu valdkonnast. Kuna isikute terviseandmeid käsitletakse tugevamat kaitset nõudvate eriliigiliste isikuandmetena, on nende teisene kasutus äärmiselt piiratud. Terviseandmete turvaline töötlemine võimaldaks lahendada tervishoius esinevaid probleemkohti. Seetõttu on PETidest teadlikkuse tõstmine ning nende edukate rakenduslugude ja parimate praktikate jagamine kriitilise tähtsusega.

Terviseandmete taaskasutus ei tohi tulla isikute eraelu kaitse arvelt ja seda positsiooni toetavad nii IKÜM, Euroopa andmestrateegia[4], Euroopa andmehalduse määrus[5] kui ka Euroopa digiturgu-

[4] Euroopa Komisjon, Euroopa andmestrateegia, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_et (viimati külastatud 01.03.2023) [9].

[5] Euroopa Parlamendi ja nõukogu määrus (EL) 2022/868, 30. mai 2022, Euroopa andmehalduse kohta ning millega muudetakse määrust (EL) 2018/1724 (andmehalduse määrus) https://data.europa.eu/eli/reg/2022/868/oj

de määrus[6]. IKÜMi lõimitud andmekaitse printsiibi järgimine nõuab asjakohaste tehniliste meetmete rakendamist kõrge riskiga andmetöötluses, sh eriliigiliste andmete nagu terviseandmete ja geeniandmete töötlemisel.[7]

Isiklike terviseandmete näitel on igaühel lihtne kujutleda sedagi, milline väärtus oleks sellel, kui andmete töötlemisel näiteks teadusuuringutes, rahvatervise meetmete kujundamisel või e-tervise rakenduste arendamisel oleksid tagatud nii andmete seostamatus, läbipaistvus kui ka andmetöötlusse sekkutavus. Korrektselt juurutatuna võiksid need oluliselt suurendada inimeste valmisolekut oma terviseandmeid teisesesse kasutusse lubada.

## 2.3 Privaatsustehnoloogiad kaitsevad ka ettevõtete andmeid

Ettevõtete või organisatsioonide andmed, sh ärisaladus, mis ei sisalda isikuandmeid, ei ole kaetud samasuguste privaatsuskaitse nõuetega nagu isikuandmed. Seega ei rakendu viidatud andmetele ei lõimitud andmekaitse põhimõte ega teised IKÜMi nõuded. Avalikus sektoris kohaldub teatud juhtudel asutuse siseseks teabeks märkimise kohustus, kui teabe avaldamine kahjustaks ärisaladust (nt avaliku teabe seadus, § 35 lg 1 p 7[8]). Ka siin on privaatsuskaitse tehnoloogiad rakendatavad. Kui need tehnoloogiad muutuvad süsteemide arenduses tavapärasemaks, siis kaasneb ka muude andmete (nt ärisaladuse) parem kaitse ning seeläbi kahanevad riskid ettevõtete andmetele. Sellel võib omakorda olla positiivne mõju ka reaalajamajanduse arengule.

Täiendava kasuna võib välja tuua, et ettevõtted, kes kasutavad privaatsuskaitse tehnoloogiaid, võivad saada klientide ja tarbijate silmis suurema usaldusväärsuse. Lisaks aitavad need tehnoloogiad ettevõtetel vähendada riske, andmelekkeid ja volitamata juurdepääse. Vastupidisel juhul võivad kaasneda soovimatud tagajärjed, mis võivad põhjustada ka reaalset majanduslikku kahju.

## 2.4 Privaatsustehnoloogiad toetavad lõimitud andmekaitset

Isikuandmete kaitse üldmääruse artikkel 25 näeb isikuandmete töötlemisel ette lõimitud ja vaikimisi andmekaitse rakendamist. Tegemist on teineteist täiendavate kontseptsioonidega, mis tõhustavad andmekaitset[9]. Privaatsuskaitse tehnoloogiate kasutamine aitab demonstreerida lõimitud ja vaikimisi andmekaitse lähenemisviisi rakendamist [10].

Lõimitud andmekaitse, mis on reguleeritud IKÜMi artikli 25 lõikes 1, tähendab, et isikuandmete vastutav töötleja peab rakendama nii töötlemisvahendite kindlaksmääramisel kui ka isikuandmete töötlemise ajal asjakohaseid tehnilisi ja korralduslikke meetmeid (IKÜM artikkel 25 lõige 1). Nende meetmete hulka kuuluvad ka privaatsuskaitse tehnoloogiad, mis aitavad muuhulgas tõhusalt rakendada andmekaitse põhimõtteid ning integreerida andmete töötlemise protsessi-

(viimati külastatud 01.03.2023) [10].

[6] Euroopa Parlamendi ja nõukogu määrus (EL) 2022/1925, 14. september 2022, mis käsitleb konkurentsile avatud ja õiglaseid turge digisektoris ning millega muudetakse direktiive (EL) 2019/1937 ja (EL) 2020/1828 (digiturgude määrus) https://data.europa.eu/eli/reg/2022/1925/oj (viimati külastatud 01.03.2023) [11].

[7] Lõimitud andmekaitse kohta vt täiendavalt IKÜM artikkel 25 ja põhjenduspunkt 78.

[8] Avaliku teabe seadus, RT I, 22.03.2011, 10, §35 https://www.riigiteataja.ee/akt/122032011010 (Viimati külastatud 01.03.2023)

[9] Euroopa Andmekaitsenõukogu, Guidelines 4/2019 on Article 25. Data Protection by Design and by Default [12]. Version 2.0. Adopted on 20 October 2020. Punkt 5. https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf (viimati külastatud 01.03.2023).

[10] Chapter 5: Privacy-enhancing technologies (PETs). Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance [13], September 2022, lk 3, https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf (viimati külastatud 01.03.2023).

desse vajalikke kaitsemeetmeid [11]. Sellised rakendatavad meetmed peavad olema tõhusad, tagama soovitud tulemusi ja vastutav töötleja peab olema suuteline ka seda tõendama (vt IKÜM, põhjenduspunkt 74 ja Euroopa Andmekaitsenõukogu suuniseid 4/2019) [12].

Vaikimisi andmekaitse on sätestatud IKÜMi artikli 25 lõikes 2. Selle kohaselt peab vastutav töötleja rakendama asjakohaseid tehnilisi ja korralduslikke meetmeid, millega tagatakse, et vaikimisi töödeldakse ainult isikuandmeid, mis on vajalikud töötlemise konkreetse eesmärgi saavutamiseks. See kehtib nii kogutud isikuandmete hulga, nende töötlemise ulatuse, säilitamise aja kui ka kättesaadavuse kohta. Nende meetmetega tagatakse eelkõige see, et vaikimisi ei tehta isikuandmeid ilma asjaomase isiku sekkumiseta määramata füüsiliste isikute ringile kättesaadavaks.

Termin "vaikimisi"(ingl k *by default*) viitab näiteks tarkvararakendusele, arvutiprogrammile või seadmele määratud konfigureeritava sätte eelnevalt olemasolevale või eelvalitud väärtusele, mida mõnikord nimetatakse ka (tehase) eelseadistusteks [13]. Andmekaitse Inspektsioon on selgitanud, et vaikimisi andmekaitse tähendab isikuandmete töötlemisel valiku tegemist seoses seadistuste väärtuste või töötlemise võimalustega, mis on määratud või ette nähtud töötlemise süsteemis. Vaikimisi võib toimuda vaid selline töötlemine, mis on rangelt vajalik seatud seadusliku eesmärgi saavutamiseks [14].

Samad nõuded kehtivad ka isikuandmete teaduslikule kasutusele ja teistele uudsetele andmete taaskasutusviisidele. Isiku õigust tema privaatsusele toetavad lisaks IKÜMile näiteks Euroopa andmestrateegia [15], digiturgude määrus [16], andmehalduse määrus[17] ja teised õigusaktid.

Seega on privaatsuskaitse tehnoloogiatel tähtis roll nii uute andmetöötlussüsteemide ehitamisel kui ka olemasolevate uuendamisel. Seda tuleb kaaluda ka olukordades, kus töödeldavad andmed on õiguslikus mõistes pseudonüümsed. Püüelda tuleb sealjuures nii seostamatuse, sekkutavuse kui läbipaistvuse poole.

## 2.5 Lõimitud ja vaikimisi andmekaitse on osa süsteemide loomulikust arengust

Peatükid 2.1, 2.2 ja 2.3 andsid privaatsuskaitse tehnoloogiate rakendamiseks sotsiaalmajanduslikku tausta ning peatükk 2.4 näitas seda põhimõtet järgima suunavad õiguslikud alused.

Nende ootuste täitumiseks tuleb ehitada uusi süsteeme ja täiendada olemasolevaid. Õiguslikest nõuetest lähtuvalt on loogiline, et selle protsessi käigus edendatakse süsteemides lõimitud ja vaikimisi andmekaitset, kahandades võimalusel töödeldavate andmete hulka ning juurutades privaatsuskaitse tehnoloogiaid.

Sellisel moel privaatsuse edendamine on ka kooskõlas andmete täiendava kasutuselevõtuga, isegi kui selle käigus mõnede andmete kasutamisest loobutakse või neid kasutatakse edasi väiksemas mahus. Kogemuste ja süsteemide küpsuse kasvades saab ka selgemaks milliste andmete töötlemisega kaasneb kasu ja mõju ning millised teenused või kratid ennast ei õigusta. Näiteks võidakse teadustöö käigus leida, et uute meetoditega on sama otsusetoe süsteemi jaoks vaja

[11] Chapter 5: Privacy-enhancing technologies (PETs) [13], lk 4.
[12] Euroopa Andmekaitsenõukogu, Guidelines 4/2019 on Article 25 [12], Punkt 13.
[13] Euroopa Andmekaitsenõukogu, Guidelines 4/2019 on Article 25 [12], Punkt 40.
[14] Andmekaitse Inspektsioon, Lõimitud ja vaikimisi andmekaitse, 06.07.2022, https://www.aki.ee/et/loimitud-ja-vaikimisi-andmekaitse (viimati külastatud 01.03.2023). Vt ka, Euroopa Andmekaitsenõukogu, Guidelines 4/2019 on Article 25 [12], Punkt 41.
[15] Euroopa Komisjon, Euroopa andmestrateegia [9].
[16] Euroopa Parlament ja Euroopa Liidu nõukogu, digiturgude määrus [11].
[17] Euroopa Parlament ja Euroopa Liidu nõukogu, andmehalduse määrus [10].

vähem sisendandmeid. Sellisel juhul on otstarbekas uuendada ärianalüüsi ja vähendada süsteemis kasutatavate andmete hulka.

**Süsteemne töö privaatsuseesmärkide saavutamise suunas on võimalik, kui kaasame digiühiskonna teenuste ja süsteemide elutsüklisse privaatsustehnika praktikad.**

# 3 Privaatsuskaitse tehnoloogiate rakendamise kontseptsioon

## 3.1 Privaatsustehnika roll süsteemide elutsüklis

Privaatsustehnika (ing k *privacy engineering*) on distsipliin, mille eesmärkideks on privaatsusriskide kahandamine, arendusressursside kulutamise otsuste põhjendamine ning kaitsemeetmete toimiv teostamine infosüsteemides. Kui privaatsustehnikat efektiivselt rakendada, on lihtsamalt võimalik täita privaatsuseesmärke ja rahuldada andmekaitsenõudeid.

Privaatsustehnika võib sisaldada erinevaid samme, näiteks nõuete, talitlusmudelite, riskide ja mõjude analüüsi, arhitektuuri loomist, infosüsteemide kavandamist ja teostamist. Seda kõike tehakse süsteemselt, lähtuvalt privaatsuseesmärkidest. Nagu arutlesime peatükis 2.5, saab selline arendustegevus olla efektiivne, kui see on kohandatud süsteemi tavapärase elutsükliga.

Süsteemi (näiteks IT-süsteemi või teenuse) elutsükkel sisaldab arendust, käitlemist, hindamist ja kasutuselt kõrvaldamist. Võime öelda, et iga süsteem on sellise elutsükli mingis etapis ning järgmised arendustegevused on võimalik selle etapi järgi tuletada. Näiteks võime eeldada, et käitluses olevat infosüsteemi mingil hetkel hinnatakse ja otsustatakse, kas seda on vaja edasi arendada või maha kanda. Igal sellisel sammul on vaja teha ka tegevusi ja otsuseid, mis aitavad süsteemiga seotud privaatsuseesmärke senisest paremini saavutada.

Ühe sobiva allikana kirjeldab privaatsustehnika rakendamist süsteemide elutsüklis tehniline aruanne ISO/IEC TR 27550 [14]. ISO/IEC TR 27550 selgitab, kuidas privaatsustehnika on seotud süsteemide arenduse, turvatehnika ja riskihaldusega. Lisaks kirjeldab viidatud tehniline aruanne, kuidas privaatsustehnikat rakendada teadmuse halduses, riskihalduses, nõuete analüüsis ja arhitektuuri kavandamises. Selle lähenemine on üldine ja sobib kasutamiseks nii tarkvara elutsüklis (näiteks ISO/IEC/IEEE 12207 [15] alusel), teenuste ja toodete elutsüklis (näiteks ISO/IEC/IEEE 29148 [16] alusel) kui ka üldisemate süsteemide elutsüklites (näiteks ISO/IEC/IEEE 15288 [17] alusel).

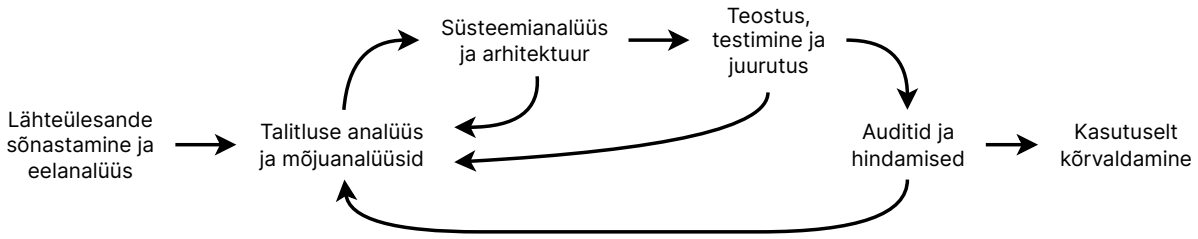
**Eelnevast lähtuvalt soovitame lõimitud ja vaikimisi andmekaitset ja privaatsuskaitse tehnoloogiaid Eesti digiühiskonnas juurutada süsteemselt, tuues privaatsustehnika tegevused tavapärasesse süsteemide arenduse protsessi.**

## 3.2 Eesti digiriigi süsteemide lihtsustatud elutsükkel

Eesti e-riigis ei ole võimalik viidata ühtsele infosüsteemide või teenuste arenduse standardile, mida järgiksid kõik asutused ühtemoodi. Siiski on nii rahastusreeglitest, ühistest mittefunktsionaalsetest nõuetest kui ka teenuseid pakkuvate organisatsioonide ülesehituse sarnasustest lähtuvalt mitmeid ühiseid osi.

Joonisel 1 toodud elutsükkel on lihtsustatud alamosa, mis on inspiratsiooni saanud teenuste loomise arenduseelsest plaanimisest, arendamiseks tehtavate hangete ülesandepüstitustest ja teenuste jätkuvast täiendamisest Eestis. Hoolimata üldistuse tasemest, on see kooskõlas süsteemide üldiste elutsüklistandarditega nagu ISO/IEC/IEEE 12207 [15] ja ISO/IEC/IEEE 15288 [17].

Tsüklilisus tähendab, et kuidagi peab olema võimalik pöörduda tagasi varasemate etappide juurde. Siin toodud lihtsustatud tsüklis võib nii arenduse kui käitlemise etappidest liikuda tagasi analüüsi juurde. Seda võivad põhjustada sisemised nõuded (näiteks uute nõuete selgumine aren-




**Joonis 1. IT-süsteemi lihtsustatud elutsükkel**

duse käigus) või väliskeskkonnast tulenevad nõuded (audit, õiguslik hinnang, võimalus süsteemi parendamisse investeerida). Elutsükkel lõpeb süsteemi kasutuselt kõrvaldamisega.

Järgmistes alajaotistes kirjeldame, kuidas privaatsustehnikat igas elutsükli etapis rakendada.

## 3.3 Privaatsustehnika rakendamine elutsükli etappides

### 3.3.1 Privaatsustehnika ülesande sõnastamisel ja eelanalüüsil

**Etapi eesmärk.** Organisatsiooni arengu käigus võivad selle liikmed leida, et edasiseks arenguks on vaja luua uus süsteem, milles töödeldakse andmeid. Kui nende andmete hulka kuuluvad mingis töötlemise etapis isikustatud või isikustatavad andmed, tuleb koheselt alustada lõimitud andmekaitsele mõtlemisega.

**Privaatsustehnika rakendamise juhised.**

1. **Loetle andmete töötlemisega seotud osapooled ja andmevood kõrgel tasemel.** Varajases arenduse faasis on nii osapoolte kui ka andmete liikumise kohta teave olemas vaid väga üldisel tasemel. Siiski on võimalik loetleda peamised andmestikud ja töödeldavad andmed, kus andmeid päritakse, nende töötlejad ja oodatavad liikumised töötlejate vahel.
2. **Sõnasta nõuded andmetöötluse minimeerimiseks.** Andmevooge ja osapooli uurides võib selguda, et mõni töötlev osapool saab oma töö ära teha väiksema andmete hulgaga või on mõnede andmete puhul tegemist eriliigiliste või muud moodi tundlike andmetega. Sellisel juhul tuleb süsteemi edasise arenduse nõuetesse lisada asjakohased korralduslikud ja tehnilised meetmed, millega minimeerida nende andmete töötlemine.

**Soovitused.** Oluline on märkida, et selles etapis ei pea veel välja valima konkreetseid privaatsuskaitse tehnoloogiaid – pigem tuleb sõnastada nõuded edasisele arendusele. Sõltuvalt süsteemi suurusest võib see etapp olla jagatud mitmeks alametapiks, mida teostavad erinevad osapooled. Näiteks võib ülesandepüstituse välja mõelda mõni riiklik asutus, kuid eelanalüüsi teostab hanke korras mõni teenuseandja. Sellisel juhul on otstarbekas koheselt järgida nõuandeid hangitavates töödes privaatsustehnika rakendamise korraldamise kohta (vt alajaotis 3.4.1).

### 3.3.2 Privaatsustehnika talitluse ja mõjude analüüsil

**Etapi eesmärk.** Talitluse analüüsi käigus täpsustatakse osapoolte ülesandeid, süsteemi funktsioone ja selleks vajalikke andmevooge. Tulemina tekivad talitluse mudelid, milles võib kirjeldada andmebaaside struktuuri ja nende kasutamise detaile. Mudelitel peab olema piisav täpsus, mis võimaldaks hinnata, milliseid andmebaase või -objekte iga osapool töötleb ning millisel isikustatavuse tasemel on töötlemine süsteemi eesmärgi täitmiseks vajalik.

Kui teatavat tüüpi isikuandmete töötlemise, eelkõige uut tehnoloogiat kasutava töötlemise tule-

musena ning isikuandmete töötlemise laadi, ulatust, konteksti ja eesmärke arvesse võttes tekib tõenäoliselt füüsiliste isikute õigustele ja vabadustele suur oht, peab vastutav töötleja enne isikuandmete töötlemist hindama kavandatavate isikuandmete töötlemise toimingute mõju isikuandmete kaitsele (IKÜM artikkel 35 lõige 1). Näiteks on eriliigiliste andmete ulatusliku töötlemise puhul andmekaitsealase mõjuhinnangu läbiviimine kohustuslik (IKÜM artikkel 35 lõige 3 punkt b). Samuti laieneb kohustus juhtudele, kus toimub isikuandmete süstemaatiline ja ulatuslik hindamine, mis põhineb automaatsel töötlemisel, sh profiilianalüüsil, ja millel põhinevad otsused, millel on andmesubjekti jaoks õiguslikud tagajärjed või muud samaväärsed mõjud (IKÜM artikkel 35 lõige 3 punkt a). Andmekaitsealane mõjuhinnang tuleb teha ka avalike alade ulatusliku süstemaatilise jälgimise korral (IKÜM artikkel 35 lõige 3 punkt c). Tegemist ei ole ammendava loeteluga olukordadest, mis nõuavad andmekaitsealase mõjuhinnangu läbiviimist.

Eesti riigisisese andmetöötluse puhul on Andmekaitse Inspektsioon määranud, et ulatusliku andmetöötlusega on tegemist juhul, kui töödeldakse:

1. eriliiki või süütegude andmeid 5 000 ja enama inimese kohta;
2. suurt ohtu põhjustavaid andmeid 10 000 ja enama inimese kohta;
3. ülejäänud isikuandmed 50 000 ja enama inimese kohta[18].

IKÜMi põhjenduspunktis 91 on välja toodud ka olukorrad, kus isikuandmete töötlemist ei tuleks lugeda ulatuslikuks. Näiteks ei peaks andmekaitsealane mõjuhinnang olema kohustuslik juhul, kui patsientide või klientide isikuandmeid töötleb üksik arst, muu tervishoiutöötaja või jurist.

Andmekaitse Inspektsioon on sedastanud, et kuigi teatud juhtudel tuleneb mõjuhinnangu teostamise nõue IKÜMist, on see andmetöötlejale ka hea tööriist, vältimaks ebasoovitavate stsenaariumide realiseerumist.[19] Seetõttu soovitame mõjuhinnangu läbi viia kõikidel juhtudel, kui kavandatakse isikuandmete töötlemist, sh kui seda tehakse uuel viisil või töödeldatakse isikuandmete uusi kategooriaid. Mõjuhinnangu läbiviimine tõstab oluliselt privaatsustehnika rakendamise efektiivsust ja privaatsuseesmärkide saavutamise tõenäosust.

**Privaatsustehnika rakendamise juhised.**

1. **Täpsusta andmevooge ja osapooli.** Täpsusta töötlejate rollid (andmesubjekt, vastutav, kaasvastutav töötleja või volitatud töötleja).
2. **Kogu andmekaitsenõuded.** Kogu osapoolte eesmärkidest lähtuvad ja süsteemi funktsionaalsust määravad nõuded. Kogu õigusruumist, lepingulistest nõuetest, standarditest ja organisatsiooni printsiipidest ja poliitikatest lähtuvad nõuded.
3. **Tuvasta osapoolte õigused ja võimekused.** Kas mõnedel osapooltel on juba olemas õiguslik alus vajalike andmete töötlemiseks? Kas on olemas privaatsustehnoloogiaid rakendavad teenused, privaatsustehnika rakendamise kogemused ja praktikad?
4. **Koosta esmane andmekaitse mõjuanalüüs.** Talitlusmudelite põhjal on juba võimalik teha esmased riski- ja mõjuanalüüsid. Analüüside tulemusena selguvad osapooled ja töötlemise tegevused, mille risk on kõrge ja mis vajavad täiendavat andmetöötluse minimeerimist või privaatsuskaitse tehnoloogiate rakendamist.

[18] Andmekaitse Inspektsioon, Isikuandmete töötleja üldjuhend (2019), ptk 5. Andmekaitsealane mõjuhinnang. https://www.aki.ee/sites/default/files/dokumendid/isikuandmete_tootleja_uldjuhend.pdf (viimati külastatud 02.03.2023), [18].

[19] Andmekaitse Inspektsioon, Isikuandmete töötleja üldjuhend (2019) [18], ptk 5. Vt ka, Artikli 29 alusel asutatud andmekaitse töörühm, Suunised, mis käsitlevad andmekaitsealast mõjuhinnangut ja selle kindlaksmääramist, kas isikuandmete töötlemise tulemusena „tekib tõenäoliselt suur oht" vastavalt määrusele (EL) 2016/679 viimati muudetud ja muudatused vastu võetud 4. oktoobril 2017, https://ec.europa.eu/newsroom/article29/items/611236 (viimati külastatud 02.03.2023), [19].

**Soovitused.** Süsteemi talitluse mudelite kaardistamiseks on mitmeid levinud tööriistu, näiteks UML tegevusskeemid või BPMN talitlusmudelid. Mõlemale on tehtud ka laiendusi, mis aitavad kaardistada osapooli, andmevooge, teha riski- ja mõjuanalüüse (nt Privacy-Enhanced BPMN [20]). On tööriistu, mille abil saab teha lekkeanalüüse ja võrrelda, millised on süsteemide garantiid, kui privaatsuskaitse tehnoloogiaid on rakendatud ja kui neid ei ole rakendatud [21].

### 3.3.3 Privaatsustehnika süsteemianalüüsi ja arhitektuuri etapis

**Etapi eesmärk.** Süsteemianalüüsi ja arhitektuuri kavandamise käigus detailitakse süsteemi mudelid selliseks, et nende põhjal saaks alustada teostust. Osapoolte ja talitluse tasemelt liigutakse tehnilise kirjelduse juurde ning kavandatakse süsteem, mis täidab nõudeid. Selles etapis tehakse ka privaatsuskaitse tehnoloogiate valikud.

**Privaatsustehnika rakendamise juhised süsteemianalüüsil.**

1. **Tuleta privaatsusnõuetest funktsionaalsed nõuded.** Kavandamise käigus tuleb leida nõuetele sobiv teostus. Näiteks, kui privaatsusnõuetes on vajadus saada andmesubjekti teavitatud nõusolek, tuleb (hiljemalt) süsteemianalüüsi etapis kavandada nõusoleku võtmise, tühistamise ning soovitatavalt ka sellega seotud läbipaistvuse ja sekkutavuse funktsioonid. Siia kategooriasse kuluvad ka näiteks logimise, säilitustähtaegade ja hävitamisega seotud nõuded.
2. **Tuleta privaatsusnõuetest mittefunktsionaalsed nõuded.** Mõned privaatsusnõuded ei mõjuta otseselt loodavat süsteemi, vaid selle keskkonda. Näiteks võib olla vajalik privaatsuse juhtimise süsteemi juurutamine organisatsiooni kvaliteedisüsteemi või teatud nõuetele vastava andmekeskuse või pilvandmetöötluse pakkuja kasutamine.
3. **Tuvasta süsteemi osad, mis vajavad andmetöötluse minimeerimist.** Kui mõjuanalüüside käigus on leitud, et mõnede andmete töötlemine mõne osapoole juures on kõrgema riskiga, tuleb süsteemianalüüsi ja arhitektuuri etapis hinnata, kas privaatsuskaitse tehnoloogiaga saab vastava töötluse riski leevendada (funktsionaalsust säilitades) või tuleb funktsionaalsust vähendada.

**Privaatsustehnika rakendamise juhised arhitektuuris.**

1. **Privaatsuskaitse tehnoloogiate valik.** Lähtuvalt süsteemianalüüsist, arhitektuurist ja mõjuanalüüsidest valitakse sobivad privaatsuskaitse tehnoloogiate kandidaadid. Kui sobivaid privaatsuskaitse tehnoloogiaid on mitu, siis võib koostada mitu kandidaatarhitektuuri ja võrrelda nende omadusi (vaata järgmist juhist).
2. **Sobivaima arhitektuuri valimine kandidaatide seast.** Süsteemi loomisel võib olla mitu arhitektuuri kandidaati (sõltumatult sellest, kas neis on erinevad privaatsuskaitse tehnoloogiad). Nende vahel valiku tegemiseks on parim viis paika panna hindamiskriteeriumid nagu keerukus, turvaeesmärkide täitmine, privaatsuseesmärkide täitmine, õigusruumiga vastavus, aga ka arenduse ja ülalpidamise hind, ja viia läbi võrdlev hindamine. Selle tulemiks on süsteemi teostamise aluseks olev arhitektuur.
3. **Privaatsusmeetmete nimekirja koostamine.** Kui arhitektuur on koostatud, siis soovitame koostada ka privaatsuse tagamise meetmete nimekirja, kus on loetletud süsteemis rakendatud korralduslikud ja tehnilised meetmed privaatsuseesmärkide saavutamiseks. Sellise nimekirja saab lisada näiteks andmekogu dokumentatsiooni hulka ning sellega tõsta süsteemi läbipaistvust, aidates kaasa privaatsuseesmärkide saavutamisele.
4. **Uuenda mõjuanalüüsi pärast nõuete ja arhitektuuri detailimist.** Kui süsteemi nõuded ja arhitektuur on täpsustatud (nt välja valitud konkreetsed tehnoloogiad ja nende pakkujad),

tuleb neile vastavalt uuendada riski- ja mõjuanalüüse, et hinnata, kas lisandunud funktsionaalsus ja meetmed on aidanud riske kahandada ja privaatsuseesmärke saavutada või on näiteks mõne uue teenusepakkuja lisamine riske tõstnud.

**Soovitused.** Standarditel ISO 9001 ja/või ISO/IEC 27001 põhineva juhtimissüsteemiga organisatsioonidel aitab privaatsuse haldamise süsteemi luua standard ISO/IEC 27701:2019 (*Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management*) [22]. Privaatsust ja andmekaitset käsitlevad moodulid on olemas ka Eesti infoturbestandardis E-ITS [23]. Maailmas kasutatakse veel ka standardimisorganisatiooni OASIS spetsifikatsiooni privaatsuse halduse etalonmudelit *Privacy Management Reference Model* (PMRM) [24], kuid see on Eestis vähem levinud.

Arhitektuuri kandidaatide vahel valikute tegemiseks on Euroopas ja Eestis edukalt rakendatud Carnegie Melloni ülikoolist pärit arhitektuuride võrlemise metoodikat *Architecture Tradeoff Analysis Method* ehk ATAM [25]. ATAM toob sisse erinevate hindamiskategooriate kaalud, millega saab süsteemi looja kas poliitika- või väärtuspõhiselt eelistada mõnda hinnatavat aspekti teisele (näiteks privaatsuseesmärkide saavutamist lihtsusele). Lisaks saab ATAMit rakendada töötoa formaadis, kogudes ja koondades erinevaid arvamusi ja luues neist ühtse hindamise tulemuse, mis võtab arvesse kõigi arvamust.

### 3.3.4 Privaatsustehnika teostamisel, testimisel ja juurutamisel

**Etapi eesmärk.** Arenduse, teostamise, testimise ja juurutuse käigus ehitatakse valmis süsteemid, liidestatakse andmebaasid, teised infosüsteemid ja teenused ning kontrollitakse süsteemi korrektsust. Valminud süsteem juurutatakse ja käivitatakse.

**Privaatsustehnika rakendamise juhis.**

1. **Järgi muudatuste juhtimisel privaatsuseesmärke ning uuenda riskianalüüsi ja privaatsusmeetmete nimekirja.** Arenduse käigus tuleb ikka ette plaanide muutuseid. Privaatsustehnika seisukohalt on kõige tähtsam jälgida, et süsteemis muudatusi tehes ei kahjustataks privaatsuseesmärke. Seega tuleb muudatusi kaaludes hinnata, kas need tõstavad riske (ja süsteem vajab seetõttu täiendavaid privaatsusmeetmeid) või langetavad riske (ja mõni meede ei ole enam vajalik).

**Soovitused.** Privaatsustehnika on ühilduv erinevate süsteemide arenduse metoodikatega (sh väle arendus). Väleda arenduse juures on privaatsuskaitse tehnoloogiate analüüsi ja teostuse tööd üldises tööjärjes. Sprindid panustavad privaatsutehnika protsesside väljunditesse (riskianalüüs, nõuete analüüs, arhitektuur, detailanalüüs). Kõik arendajad ei pea olema privaatsustehnika asjatundjad, piisab ühest. Arendusmeeskonna ühel liikmel võib olla eraldi roll, mis vastutab andmekaitsega seotud tehniliste valikute eest (privaatsuse tooteomanik).

### 3.3.5 Privaatsustehnika auditi ja hindamise käigus

**Etapi eesmärk.** Süsteemi auditid ja hindamised võivad toimuda mitmel põhjusel. Näiteks võib tegemist olla regulaarse turvaauditiga. Samuti võib süsteemi rakendav asutus põhjalikumalt analüüsida oma juhtimisala tööd ja süsteemide vajalikkust.

Iga sellise kontrolli ülesandepüstitus on erinev. Näiteks kontrollitakse, kas süsteem vastab nõuetele ning hinnatakse, milliseid osi on vaja uuendada või täiendada. Selles etapis on võimalik teha otsus lõimprivaatsuse ja privaatsuskaitse tehnoloogiate lisamiseks süsteemidele, mis neid seni ei kasutanud. Äärmisel juhul, kui süsteemi loomise põhjustanud lähteülesanne enam lahendamist

ei vaja, võidakse süsteem ka käitlusest eemaldada.

**Privaatsustehnika rakendamise juhised.**

1. **Kontrolli privaatsuseesmärkide täitmist.** Kui süsteemi arenduses ei ole varem privaatsustehnikat rakendatud, on süsteemi auditeerimine või suuremate arenduste planeerimine selleks hea koht. Esimeseks sammuks on teadvustada, et soovitakse lõimitud andmekaitse rakendamist rakendada. Teine samm on vaadata üle süsteemi hetkeseis (näiteks teostatud mõjuanalüüsid) ning hinnata, milliste privaatsuseesmärkide suunas järgmises arendustsüklis töötama peaks. Kolmas eesmärk on leida ressursid privaatsustehnika arendamiseks ja/või rakendamiseks.
2. **Toeta protsessi seniste privaatsustehnika tegevustega.** Kui privaatsustehnikat on juba rakendatud, siis selle rakendamise käigus kogutud dokumentatsioon toetab nii auditeid kui hinnanguid. Privaatsusmeetmete nimekiri toetab turvaauditeid. Privaatsusmõjude analüüs aitab hinnata privaatsuseesmärkide täitmist.

### 3.3.6 Privaatsustehnika süsteemi kasutuselt kõrvaldamisel

**Etapi eesmärk.** Vahel selgub, et süsteem on oma töö ära teinud, selle käitamiseks ei ole organisatsioonil enam ressursse või on see moraalselt vananenud ning asendatakse uue ja sobivamaga. Sellistel juhtudel lõpetatakse vana süsteemi kasutamine.

**Privaatsustehnika rakendamise juhised.**

1. **Veendu, et süsteemi kasutuselt kõrvaldamine ei mõju halvasti organisatsiooni privaatsuseesmärkide täitmisele.** Mõni infosüsteem toetab läbipaistvust või sekkutavust, hallates andmesubjekti nõusolekuid, näidates talle tema andmete töötlemise ulatust või lubades selle kohta juhiseid anda. Teine süsteem aitab andmetöötlust minimeerida, juurutades privaatsuskaitse tehnoloogiaid, mis vähendavad andmete seostatavust andmesubjektidega. Selliste süsteemide kaotamine võib organisatsiooni privaatsuseesmärkidele olla halva mõjuga, kui süsteemi ei asendata, ning siis tuleks kaaluda, kas kasutuselt kõrvaldamine on ainus võimalus.
2. **Uuenda külgnevate süsteemide riskianalüüse.** Kui süsteem lõpetab töö, võib muutuda sellega seotud süsteemide privaatsusmõju. Seda näiteks juhul, kui need süsteemid ei pea enam saatma andmeid eemaldatud süsteemi.

## 3.4 Etapist sõltumatud soovitused

### 3.4.1 Privaatsustehnika rakendamise korraldamine hangitavates töödes

**Selgitus.** Süsteemi rajav organisatsioon võib kõikides elutsükli etappides korraldada hankeid tööde tegemiseks. Järgmised soovitused aitavad sellistes projektides privaatsustehnikat rakendada.

Üldise reeglina – kui hanke järgi süsteemi või selle osade tarnija peab tegema tehnilisi valikuid, sh privaatsuskaitse tehnoloogiate valikuid, siis peaks hankija ette andma nõuded, millistes etappides tuleb andmetöötlust vähendada või vältida. Näiteks, kui vastutav töötleja soovib täielikult vältida isikustavate andmete töötlust ning rakendab selle eesmärgi saavutamiseks privaatsuskaitse tehnoloogiaid. Või juhul, kui tegemist on andmete avaldamisega, tuleks ette anda nõuded, millisel tasemel taasisikustamist peaks tehniliselt vältima. Näiteks ”taasisikustamine ei tohi olla võimalik andmestiku töötlemisel kõrvalise infota või kõrvalist infot kasutades”. Viimasel juhul tu-

leb mõelda, kuidas välistatakse kõrvalise info kasutamine (mis ei ole pikas perspektiivis realistlik eeldus).

**Privaatsustehnika rakendamise juhised.**

1. **Lisa privaatsusmeetmed ja vajadusel privaatsuskaitse tehnoloogiad hanke lähteülesandesse.** Privaatsustehnika edukaks rakendamiseks peavad hanke kirjelduses olema vastava arendusetapi jaoks sobivad nõuded. Hanke edukuse üheks tingimuseks on ka hankija ja pakkuja kohustustes kokku leppimine. Näiteks võib vastutav töötleja hankida teatud viisil (nt privaatsuskaitse tehnoloogiatega) töötlemist mõnelt volitatud töötlejalt või kaasvastutavalt töötlejalt. Samuti võib vastutav või volitatud töötleja hankida analüüsi- või arendustöid, mis võivad vajada privaatsuskaitse tehnoloogiate juurutamist või otsest rakendamist. Hankija ei pea ise määrama tehnoloogiat, kuid võib nõuda, et minimeeritakse teatud andmete töötlemist teatud osapoole juures. Asjakohased privaatsusmeetmed võivad sõltuda näiteks õiguslikust keskkonnast, kus ettevõte tegutseb, erinevatest lepingulistest kohustustest, aga ka ettevõtte eemärkidest ja privaatsuspoliitikatest (vt nt ISO 27550 lk 18, privaatsuse valitsemise raam).
2. **Nõua pakkujalt privaatsustehnilise kompetentsi tõendamist.** Kui hanke käigus on ette näha mõne privaatsuskaitse tehnoloogia või privaatsustehnika meetodi rakendamine, siis tuleb vastavate kompetentside tõendamine panna hankes kvalifitseerumise tingimustesse.
3. **Vormista hankelepingus õiguslikud alused tarnija tööks.** Kui tarnija peab oma töö käigus töötlema hankija edastatud andmeid, sh isikustatavaid andmeid, peab hankija vastavad õigused ja kohustused lisama hankelepingusse.
4. **Kogu teavet hangitavate süsteemide privaatsusmeetmete kohta.** Kui hangitakse tarkvara või IT-teenuseid, kohusta tarnijat selgitama välja ning esitama pakutavate tehnoloogiate privaatsusmeetmed ja kasutatud privaatsuskaitse tehnoloogiad.
5. **Kogu teavet hangitavate süsteemide tarneahela kohta.** Kui tarnitav lahendus sõltub kolmandate poolte süsteemidest, kohusta tarnijat välja selgitama ja esitama teavet kõigi volitatud alamtöötlejate kohta, kes tarnitava süsteemi käitlemisel isikustatavaid andmeid töödelda võiksid. Kohusta tarnijat välja selgitama, millistes jurisdiktsioonides alamtöötlejad andmeid töötlevad ja millised on sellest lähtuvad riskid.
6. **Fikseeri vastutus jääkriskide eest.** Mõne privaatsuskaitse tehnoloogia puhul tekivad jääkriskid (nt teenuseandja, mõni tema alltöövõtja või volitatud alamtöötleja taasisikustab mõned andmed või rikub nende kasutuse reegleid). Vajalikud kohustused nende ohtude realiseerumise tõenäosuste vähendamiseks tuleb panna hanke ja lepingu tingimustesse, et pakkujad saaksid nendega arvestada.
7. **Fikseeri infoturvasündmustest ja intsidentidest teatamise kord ja kontaktisikud.** Hankija ja pakkuja vahel tuleb kokku leppida kontaktisikud, keda ja millises ajaraamis teavitatakse infoturvasündmustest ja intsidentidest. Hea oleks defineerida ka selliste olukordade mõisted, et vältida arusaamatusi.

**Soovitused.** Põhjalikumaid juhiseid hangetes privaatsustehnika rakendamise kohta annab standard ISO/IEC 27550:2019 [14].

### 3.4.2 Üldised soovitused privaatsusmõjude ja riskide analüüsiks

**Selgitus.** Eelnevates juhistes rõhutasime mõju- ja riskianalüüside tähtsust süsteemi elutsükli kõikides etappides. Selle jaoks on mitmeid metoodikaid. Üldine töövoog on järgmine.

**Privaatsustehnika rakendamise juhised.**

1. **Vali sobiv mõjude ja riskide analüüsi metoodika.** Standardimisasutused, andmekaitseasutused ja teadlased on välja töötanud mitmeid lähenemisviise ja metoodikaid privaatsusmõjude hindamiseks ja riskide analüüsiks. Nende juurutamine ei pea tähendama kogu standardi juurutamist. Efekt saavutatakse ka siis, kui võetakse üle üldine protsess ja peamised elemendid.
2. **Tuvasta võimalikud ohud ja ründed.** Andmekaitseõigus seab kohustused ja määrab trahviõiguse isikustatavaid andmeid töötlevatele organisatsioonidele. Siiski tuleb privaatsusmõjude hindamisel eraldi käsitleda ohte isikutele kui võimalikele lõppkannatajatele. Analüüsi käigus tuleb kaardistada nii mittevaralisi kui varalisi kahjusid tekitavad ohud ning määrata ründaja profiil, kelle eest isikuandmeid kaitsta soovitakse.
3. **Hinda rünnete võimalikkust lähtuvalt ajakohasest teadmisest.** Süsteemi arengu jooksul võib olla vajadus hakata arvestama varasemast võimekama ründajaga, kellel on (näiteks tehnoloogia arengu tõttu) parem võimekus andmete taasisikustamiseks. Sel juhul tuleb analüüsi vastavalt uuendada. Teavet rünnete võimalikkusest on otstarbekas hallata organisatsiooni teadmuse baasis (vt peatükk 3.4.3).
4. **Erista tõsisemad ohud, millele on kaitsemeetmeid vaja.** Tuleta rünnete võimalikkuse ja mõjutatud isikute hulga põhjal tõsisemad ohud, mille puhul on lisameetmete kasutamine vajalik.
5. **Dokumenteeri tulem.** Mõju- ja riskianalüüsid tuleb koos lähteandmetega dokumenteerida, et neid kasutada elutsükli järgmistes etappides.

**Soovitused.** Privaatsusohtude modelleerimiseks on välja töötatud metoodika nimega LINDDUN [26]. LINDDUN annab struktuuri privaatsusohtude üle arutlemiseks ja sõnastab kategooriad, mille kaudu ohtu kirjeldada. Alustajate ja väikesemate süsteemide jaoks on LINDDUNist tehtud ka lihtsustatud versioon LINDDUN GO.

Üks privaatsusmõju analüüside standard on ISO/IEC 29134:2017 (*Privacy impact assessment – guidelines*) [27]. Teine on NIST SP 800-53A Rev. 5 (*Assessing Security and Privacy Controls in Information Systems and Organizations*) [28]. Euroopa kontekstis on sobilik kasutada ka Prantsusmaa andmekaitseorganisatsiooni CNIL tööriistu, malle ja materjale [20].

### 3.4.3 Privaatsustehnika teadmuse kogumine ja haldus

**Selgitus.** Privaatsustehnika juurutajal või lahenduste teostajal on soovitav koguda ja hallata teavet, mis aitab tal teenuseid ja süsteeme ehitada. See aitab organisatsiooni privaatsusmeetmete alast küpsust tõsta ning säilitada ka juhul, kui süsteemidega töötavad inimesed vahetuvad.

**Privaatsustehnika rakendamise juhised.**

1. **Dokumenteeri privaatsustehnika tegevuste käigus kogutud materjalid ja toodetud tulemid.** Organisatsioonil on privaatsuseesmärkide kestliku täitmise jaoks otstarbekas säilitada tehtud mõjuanalüüsid, riskianalüüsid, õiguslikud analüüsid ning pretsedendid, tehnoloogiate ja teenuste kirjeldused, garantiidokumendid ja sertifikaadid. Lisaks formaalsetele materjalidele on kasu ka vähemametlikust teabest nagu teiste kasutajate kogemuslood, mis võivad inspireerida uusi rakendusi ja arhitektuure.

[20] Privacy Impact Assessment (PIA). Internetis kättesaadav: https://www.cnil.fr/en/privacy-impact-assessment-pia (viimati külastatud 02.03.2023).

### 3.4.4 Privaatsustehnika seosed info- ja küberturbega

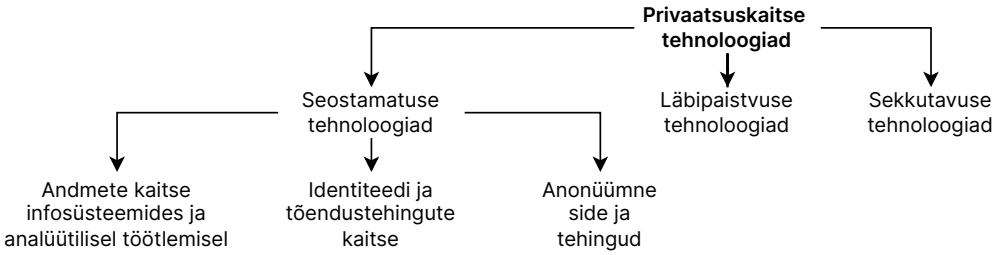
Privaatsustehnika protsessid ja tegevused on teostatavad samas taktis info- ja küberturbe tegevustega. Siiski on oluline mõista, et info- ja küberturbe juhtimine ei taga automaatselt privaatsuseesmärkide saavutamist. Näiteks läbipaistvus ja sekkutavus ei ole turvanõuded. Mõned turvatehnoloogiad aitavad saavutada seostamatust, kuid ei aita saavutada minimeerimist. Näiteks pääsuõiguste tagamine ja salvestatud andmete krüpteerimine ei tarvitse saavutada andmetöötluse minimeerimist.

Privaatsustehnika rakendamine, näiteks privaatsuse juhtimissüsteemi raames, sobib samas hästi kokku organisatsiooni infoturbe juhtimissüsteemiga. Näiteks standard ISO/IEC 27701:2019 [22] kirjeldab, kuidas standardile ISO/IEC 27001 [29] vastavas juhtimissüsteemis juurutada privaatsuse juhtimist. Ka Eesti infoturbestandardis E-ITS on andmekaitse kontseptsioonid rõhutatud.

# 4 Privaatsuskaitse tehnoloogiad

## 4.1 Privaatsuskaitse tehnoloogiate kategooriad

Privaatsuskaitse tehnoloogiate ülevaade on jaotatud kahe kategooria järgi (Joonis 2).




**Joonis 2. Privaatsuskaitse tehnoloogiate jaotus peatükkidesse**

Esimeses jaotuses on privaatsuseesmärgid, mida tehnoloogiad aitavad saavutada. Suur enamus tänaseid juurutamiskõlbulikke privaatsuskaitse tehnoloogiaid taotleb seostamatust ehk püüab vältida töödeldavate andmete isikustamist anonüümimise ja pseudonüümimise abil. Sellesse kategooriasse kuuluvate küpsete tehnoloogiliste lahenduste suur hulk ja variatsioon nõuab detailsemat liigitust. Seostamatuse tehnoloogiaid on märkimisväärselt rohkem ning need jagame kolme eraldi alajaotisesse selle alusel, millise töötlemise käigus nad isikustamist väldivad.

1. Andmetöötluse privaatsust kaitsevad tehnoloogiad (alajaotis 4.2) vähendavat privaatsete andmete leket sisend- või väljundandmetest või töötlemise protsessi käigus.
2. Identiteedi privaatsust kaitsevad tehnoloogiad (alajaotis 4.3) tõestavad kuuluvust volitatud rühma või siis isiku üksikuid tunnuseid (nt vanuse tõestamine ilma tervet isikukoodi või sünnipäeva avaldamata).
3. Kommunikatsiooni privaatsust kaitsevad tehnoloogiad (alajaotis 4.4) tagavad sõnumisaladuse kaitse ja loovad anonüümse side võimalusi (nt vilepuhujate sidelahendused).

Läbipaistvuse (alajaotis 4.5) ja sekkutavuse (alajaotis 4.6) sisuks on andmetöötluse kohta ammendava info tagamine ja (andmesubjekti) võimalus andmetöötlus katkestada. Nende privaatsuseesmärkide puhul on keerukam eristada tehnoloogiat ja selle rakendust, kuna rakendatav tehnoloogia ei tarvitse olla privaatsusspetsiifiline vaid universaalsem (nt logimine). Lisaks eeldab nende eesmärkide saavutamine ka seda, et organisatsiooni talitlus viiakse kooskõlla privaatsuspõhimõtetega. Seega on nende rakendamine tähtis juba süsteemi loomise varases etapis ning on seega sarnane andmetöötluse minimeerimise printsiibi rakendamisega (vt peatükk 3).

## 4.2 Andmete kaitse infosüsteemides ja analüütilisel töötlemisel

### 4.2.1 Pseudonüümimine

**Lihtsustatult:** Pseudonüümimine asendab otseselt isikustavad osad andmetes kaudselt isikustavatega.

**Ülevaade ja rakendamine.** Pseudonüümimine on andmestikus otseselt identifitseerivate tunnuste asendamine pseudonüümidega. Need pseudonüümid luuakse algoritmiliselt ning neid on võimalik algsete andmetega siduda ainult lisainfo olemasolul. Seostamiseks vajalik teave või algoritmid hoitakse pseudonüümitud andmetest eraldi. Kui kaks andmestikku on pseudonüümitud samade reeglite alusel ja sama võtmega, võib pseudonüümide järgi olla võimalik andmestikke ühendada.

Oluline on mõista, et pseudonüümimisena tõlgendavad juristid ja andmeteadlased vahel erinevaid asju. IKÜMi mõistes saavutavad kõik tehnoloogiad, millest on võimalik andmeid mõistliku pingutusega taasisikustada, pseudonüümimise. Pseudonüümide loomiseks on mitmeid tehnikaid ja siin alajaotises keskendume konkreetsetele tehnilistele viisidele, mis asendavad või eemaldavad otseselt isikustavaid tunnuseid.

Pseudonüümimine on üks lihtsamatest privaatsuskaitse tehnoloogiatest. Seetõttu on ta ka väga paljudes kohtades kasutusel. Pseudonüümimiseks selgitatakse välja andmesubjekti otseselt tuvastavad tunnused (n isikukood, nimi, täpne aadress) ning igale subjektile määratakse unikaalne kood ehk pseudonüüm. Otseselt tuvastavad tunnused eemaldatakse või üldistatakse (täpne aadress muudetakse kohaliku omavalitsuse, maakonna või riigi täpsusega elukohaks). Tavaliselt jäävad pseudonüümitud andmestikku alles kõik algses valimis olnud isikud.

Erinevaid andmestikke on võimalik linkida kolmanda osapoole abil, kellele antakse seosed unikaalse tunnuse (identifikaatori) ja sellele vastavate pseudonüümide vahel. Usaldatud osapool, kelleks võib näiteks olla üks andmete vastutavatest töötlejatest, loob unikaalse identifikaatori põhjal seose erinevate pseudonüümide vahel ning väljastab ainult selle seose. Selle järgi saab andmete töötleja andmestikud omavahel siduda.

Pseudonüümide genereerimise võimalusi on palju: lihtsamaid ja keerulisemaid. Meetodi valik sõltub pseudonüümimse eesmärgist ja andmestiku suurusest. Kõige lihtsam meetod on loenduri-meetod, mille puhul asendatakse ID järjekorranumbriga alates sobivast arvust $t$. Sellisel juhul puudub loodud pseudonüümi seos algse identifikaatoriga, kuid võib lekkida algse andmestiku kirjete järjekorra kohta (näiteks, kui andmestik on järjestatud perenime või isikukoodi järgi). Kui andmestikus on palju andmeid (mobiilsus, energiatarbimine), ei pruugi see meetod olla kestlik, sest vaja on alles hoida vastavustabelit, mis suurte andmemahtude korral ei ole enam korralikult hallatav. Kui pseudonüümid luua juhuslikult (kasutades näiteks pseudojuhuarvude generaatorit) kaob semantiline seos algse kirjete järjekorraga, aga ka sellel juhul on vaja hoida vastavustabelit ning juhuarvude genereerimisega lisandub vajadus tegeleda kollisioonidega ehk olukorraga, kus juhuslikult luuakse sama arv mitu korda.

Pseudonüümide loomiseks on võimalik kasutada ka näiteks krüptograafilist räsifunktsiooni[21] või krüpteerimist. Need meetodid on keerulisemad ning vajavad põhjalikumat juurutust, aga olenevalt olukorrast võib see mõistlik olla. Nendel juhtudel ei ole vaja hoida vastavustabelit algse identifikaatoriga, sest pseudonüüm on algoritmiliselt loodud identifikaatorit ja salajast võtit ka-

[21] Krüptograafiline räsifunktsioon saab sisendiks salajase võtme ja väärtuse (pseudonüümimisel identifikaatori) ning väljastab räsi (pseudonüümi).

sutades.

Võimalik on kasutada ka ilma võtmeta räsifunktsiooni, aga kui identifikaatori domeen on struktuurne ja piiratud (näiteks isikukoodid), on räsid igaühel võimalik jõuründega (läbiproovimise teel) arvutada. Teades isikukoodi koostamise reegleid ja seda, mis räsifunktsiooni on kasutatud, on see kõigi Eesti isikukoodide puhul tehtav vähem kui ühe päevaga, millest enamuse võtab vastava andmetöötlusülesande programmeerimine. Kui räsifunktsioon ei ole teada, võtab kõigepealt aega räsifunktsiooni väljauurimine, aga ka see on pigem paari päeva küsimus. Krüptograafiline räsifunktsioon kaitseb selle ründe eest, sest siis peab ründaja jõurünnet teostama kõikide võimalike salajaste võtmete kohta ja see taandub juba vastava krüptosüsteemi murdmise ülesandele, mis on tõestatult raske.

Pseudonüümida saab mitmesuguseid andmeid, näiteks teksti[22],[23] pilte ja videomaterjali[24]. Selline töötlemine võib saavutada väljundina teksti, pildi või video, mida ei saa isikustada, kuid sellist garantiid ei saa rikkaliku meediumi nagu tekst, kõne, pilt või video puhul anda.

**Turvagarantiid ja jääkriskid.** Ainult pseudonüümimine ei taga andmesubjektide privaatsust, sest Internetist paljude inimeste kohta kergesti kättesaadav lisainfo võimaldab inimesi suure tõenäosusega lihtsalt taasisikustada ilma pseudonüümimiseks kasutatud algoritme ja võtmeid teadmata. IKÜM-i mõistes on need andmed taasisikustatavad, seega kehtivad pseudonüümitud andmete töötlemisele samad nõuded nagu isikustatud andmete töötlemisele.

Pseudonüümitud andmekogu ründaja eesmärgiks on tavaliselt ühe või mitme kirje taasisikustamine. Lihtsam rünne on aga pseudonüümi taga oleva isiku kohta vähemalt ühe omaduse tuvastamine (”kõigi aines osalenud naiste eksamitulemused olid kõrged”). See ei pruugi lõppeda taasisikustamisega, aga võib olla piisav, et isikut mingil moel diskrimineerida. Näiteks võtame juhu, kus aines käis kümme inimest, kellest kaks olid naised ja mõlema naise eksamitulemused olid kõrged (95 ja 92). Meeste eksamitulemused olid ka üldiselt kõrged, aga oli kaks meest, kelle tulemused olid madalad (65 ja 58). Kui eksamitulemused on avaldatud ilma identifikaatorita (andmestikus on ainult sugu ja hinne) ja tööandjale on teada, et kandidaadid käisid selles aines, siis on kandidaatidest võimalik eelistada naisi, sest on teada, et mõlema naise eksamitulemus oli kõrge, kuigi leidus ka mehi, kelle eksamitulemused olid sama head.

**Juhised rakendajale.** Pseudonüümimise rakendaja peab esimese sammuna teadvustama meetodi piiranguid – et tulemina saadud andmetele tuleb tagada sama kaitstuse tase kui isikustatud andmetele. Kui see on vastuvõetav, siis järgmine samm on otsustada, millise meetodiga pseudonüümitakse. Üks võimalus on juurutada pseudonüümimine globaalselt. Näiteks organisatsioonisiseselt pseudonüümimise süsteemi puhul asendatakse kõikides infosüsteemides isikustatud tunnused sama pseudonüümiga.

Pseudonüümitud andmete jagamisega kaasnevad tavaliselt korralduslikud meetmed, näiteks konfidentsiaalsuslepingud. Kui asutus on andmete vastutav töötleja, on allkirjastavad andmetega töötajad tavaliselt konfidentsiaalsuslepingu. On selge, et asutuse põhitoimingud vajavad, et töötajatel oleks juurdepääs isikustatud andmetele (haiglad, Tervisekassa, politsei, koolid). Aga ka selliste asutuste sees on statistika tegemiseks soovitav kasutada pseudonüümitud andmeid, sest see vastab töötluse minimeerimise nõudele (isikukood ja nimi ei ole statistika mõttes olulised tunnused). Näiteks on võimalik tööd teha operatiivbaasi peal ja statistika jaoks luua pseudonüümitud andmeladu.

Teadusuuringute tegemisel kasutatakse ka pseudonüümimist, sest siis jääb andmestikku alles

[22] Private AI. https://www.private-ai.com (viimati külastatud 27.02.2023).
[23] Teksti anonüümija. https://github.com/orgs/buerokratt/projects/25 (viimati külastatud 27.02.2023).
[24] Google Magritte. https://github.com/google/magritte (viimati külastatud 27.02.2023).

võimalikult palju isikuid. Sellisel juhul (eriti tervise- ja sotsiaalandmete puhul) eelneb andmeväljastusele tavaliselt eetikakomitee hinnangu taotlus, kus on ära põhjendatud, miks andmeid analüüsitakse, ning kes andmeid näevad. Kui eetikakomitee on nõusoleku andnud, allkirjastatakse konfidentsiaalsusleping, kus teadlased lubavad, et nad ei otsi pseudonüümitud andmestikust üksikisiku andmeid välja ning ei ürita andmestikku osaliselt või täielikult taasisikustada. Ka siin on konfidentsiaalsusleping lisameede, millega leevendatakse taasisikustamise riski.

Avaandmete puhul sellisel tasemel korralduslikke meetmeid kasutada ei saa. Pseudonüümimine ei ole sobilik vahend avaandmete loomiseks, sest ei ole võimalik tagada, et keegi ei suuda pseudonüüme hiljem taasisikustada.

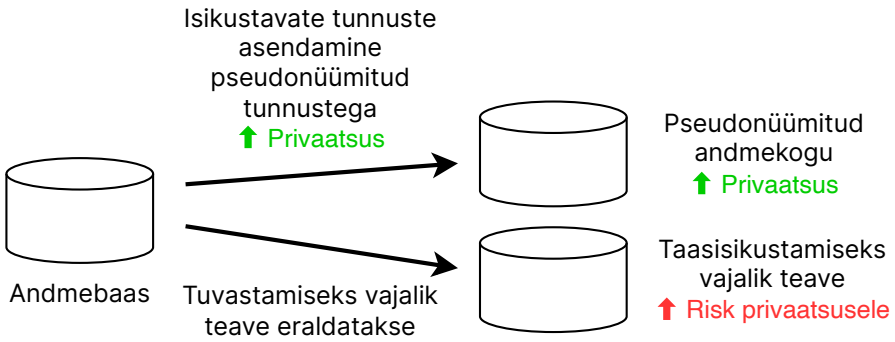
Pseudonüümimise kohta on juhendmaterjale avaldanud ENISA [30, 31].

**Õiguslikud aspektid.** Pseudonüümitud andmeid, mida saaks füüsilise isikuga seostada täiendava teabe abil, tuleks käsitada teabena tuvastatava füüsilise isiku kohta. Füüsilise isiku tuvastatavuse kindlakstegemisel tuleks arvesse võtta kõiki vahendeid, mida vastutav töötleja või keegi muu võib füüsilise isiku otseseks või kaudseks tuvastamiseks mõistliku tõenäosusega kasutada, nt teiste hulgast esiletoomine. Selleks, et teha kindlaks, kas füüsilise isiku tuvastamiseks võetakse mõistliku tõenäosusega meetmeid, tuleks arvestada kõiki objektiivseid tegureid, nt tuvastamise maksumus ja selleks vajalik aeg, võttes arvesse nii andmete töötlemise ajal kättesaadavat tehnoloogiat kui ka tehnoloogilisi arenguid. (IKÜM, pp 26)

IKÜMi pp 29 kohaselt selleks, et luua stiimuleid pseudonüümimise kasutamiseks isikuandmete töötlemisel, peaks samal vastutaval töötlejal olema võimalik kasutada pseudonüümimismeetmeid, mis võimaldavad samal ajal üldist analüüsi, kui vastutav töötleja on võtnud tehnilised ja korralduslikud meetmed, mis on vajalikud, et tagada asjaomase andmetöötluse tegemisel käesoleva määruse rakendamine, ning sellise täiendava teabe, mis võimaldab isikuandmeid seostada konkreetse andmesubjektiga, eraldi hoidmise. Andmeid töötlev vastutav töötleja peaks tähendama sama vastutava töötleja volitatud isikuid.

| **Pseudonüümimine** | **ANDMED** |
|---|---|
| **Inglise keeles:** pseudonymisation | |
| **Lühidalt:** Pseudonüümimine asendab otseselt isikustavad osad andmetes kaudselt isikustavatega. | **Arenduse keerukus:** madal |
| | **Ülalpidamise keerukus:** madal |
| | **Täpsus:** kadudega (isikustavaid väärtuseid eemaldatakse) |
| | **Privaatsusgarantii:** tõestatavat privaatsusgarantiid ei ole |
| | **Tehnoloogia küpsus:** kõrge |

**Ülevaatlik mudel:**




**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: pseudonüümimise algoritmis on juhuslik komponent.
2. Turvaeeldus: juurdepääs taasisikustamist võimaldavale teabele on tugevalt piiratud.
3. Turvaeeldus: pseudonüümitud andmestiku töötleja tegevus on piiratud.
4. Jääkrisk: ka siis kui kõik see on tehtud, on taasisikustamise risk kõrge.

**Rakendusvõimalused:**

1. Andmete avaldamine teadustööks (väga kõrge riskiga)
2. Andmete linkimise teenus (väga kõrge riskiga)
3. Asutusesiseste andmeladude loomine statistika ja uuringute jaoks

**Õiguspraktika:**

1. Pseodonüümitud andmeid käsitletakse andmekaitse õiguses kui isikuandmeid.

**Tuntumad rakendused:**

1. X-tee kodeerimiskeskus
2. Sotsiaalministeeriumi haldusala kodeerimiskeskus
3. Tartu Ülikooli Eesti Geenivaramu kodeerimiskeskus

### 4.2.2 Anonüümimine

**Lihtsustatult.** Anonüümimine on protsess, mille kaudu isikustatav teave muudetakse mitteisikustavaks.

**Ülevaade ja rakendamine.** Anonüümimine on protsess, millega isikutuvastusteave (isikukood, nimi, muu komplekt tunnuseid, mis määravad isiku üheselt) eemaldatakse andmestikust või muudetakse nii, et isikut ei saa mitte keegi otse ega kaudselt tuvastada (isegi mitte andmete vastutav töötleja ega anonüümija ise). Seega, kui kuskil (ükskõik kelle käes) on võti või tabel, mille abil on võimalik andmestik taasisikustada, siis on tegemist pseudonüümimise mitte anonüümimisega. Kui andmestikust eemaldada isikukood, nimi ja aadress, ning anda igale isikule juhuslik kood (näiteks unikaalne ID uues andmebaasitabelis), siis ei ole tegemist anonüümimisega.

Andmed on mõne osapoole jaoks anonüümsed (või anonüümitud), kui see osapool ei suuda mõistliku pingutusega andmetest ühtegi isikut tuvastada. Anonüümimise puhul ei tohi olla võimalik luua seost algse andmebaasikirje ja anonüümitud andmebaasikirje vahel. Kui taasisikustamine ei ole võimalik, ei ole enam tegemist isikuandmetega ja seetõttu ei kuulu selliste andmete töötlemine ka IKÜMi kohaldamisalasse. Seda, kas taasisikustamine on võimalik või kui anonüümne on konkreetne andmeobjekt, saab hinnata ainult juhtumipõhiselt.

Võib esineda ka olukordasid, kus vastutav töötleja avaldab või edastab kolmandale isikule andmed, mis on töödeldud selliseks, et need on kolmandate isikute jaoks anonüümsed, st et kolmandatel isikutel ei ole võimalik andmesubjekti tuvastada. Kui vastutavale töötlejale jääb see võimalus alles, siis tema vaatest on tegemist endiselt isikuandmetega. Sellises olukorras peab vastutav töötleja tagama, et ta järgib kõiki isikuandmete kaitse reegleid ja et kolmandatel isikutel ei ole tõepoolest võimalik andmeid taasisikustada.

Anonüümimise protsessis ei piisa sellest, et lihtsalt isiku identifikaatorid eemaldada, kuna kõigil isikutel on olemas kvaasi-identifikaatorid. Kvaasi-identifikaatorid on tunnused, mille abil on võimalik isik tuvastada, kui neid koos vaadata. Tunnuseid, mis komplektina võivad moodustada kvaasi-identifikaatorid, on väga keeruline andmestikus kindlaks määrata ning need sõltuvad mite ainult tunnustest vaid ka andmetest konkreetses andmestikus. Näiteks võivad sugu ja haridustase väikese asula andmestikus olla kvaasi-identifikaatorid, samas kui mõne suure linna andmestikus ei ole nende tunnuste järgi võimalik kedagi üheselt määrata. Aga kui andmestik sisaldab ka näiteks ülikooli nime, eriala ja lõpetamise aastat, võivad need tunnused kõrgemates õppeastmetes jälle moodustada kvaasi-identifikaatorid.

Kuna anonüümimise protsess on lõplik ning ei tohi olla pööratav, tuleb arvestada, et selle tehnoloogia kasutamise puhul ei saa isikuid neid puudutavatest leidudest teavitada. Näiteks on võimalik rääkida üldiselt, et sellise profiiliga inimestel on oht sattuda haiglasse järgmise poole aasta jooksul, aga pole võimalik isikut otse informeerida, et ta uuringutele läheks.

Anonüümimise meetodid jagunevad kaheks: juhuslikkuse lisamine ja üldistamine. Võimalik on kasutada mitut meetodit, näiteks üldistamist, täpsuse vähendamist ja kategoriseerimist andmestiku ettevalmistamiseks ja siis agregeerimist k-anonüümsuse saavutamiseks. Vahel piirdutakse ka ettevalmistavate meetoditega, aga sellisel juhul on taasisikustamise risk kõrge.

Juhuslikkuse lisamise alla kuuluvad näiteks müra lisamine ja tunnuse sees väärtuste permuteerimine. Mõlemad nimetatud meetodid on pigem täiendavad mitte piisavad anonüümsuse saavutamiseks. Müra lisamine muudab tunnuse väärtuseid nii, et need on vähem täpsed (näiteks muuta iga inimese kaalu $+/-5$ kg). Selle protsessi jooksul üritatakse säilitada üldist tunnuse jaotust. Lisatud müra hulk sõltub sellest, kui oluline on see tunnus analüüsis ning kui suur on mõju inimese privaatsusele, kui see tunnus avalikustatakse. Kui lisada semantiliselt ebarealistlik kogus müra, siis seda on võimalik välja filtreerida (näiteks kui pikkus meetrites, aga müra lisatakse nii nagu baasis oleks pikkuse andmed sentimeetrites).

Tunnuse sees väärtuste permuteerimise puhul seotakse väärtused erinevate inimestega. Sellise väärtuste segamise puhul jääb alles täpsed tunnuste jaotused ning vahemikud, kahjuks aga kaovad tunnustevahelised seosed (näiteks haiglasse kirjutamise põhjus, sümptomid ja osakond) või korrelatsioonid (näiteks kaalu ja pikkuse vaheline korrelatsioon). Nendele korrelatsioonidele tuginedes saab ründaja kindlaks teha, millised tunnused on permuteeritud ning osaliselt taastada algse järjestuse. Võimalik on väärtusi permuteerida nii, et seosed säiliksid, aga see on keerulisem ning võib suurendada taasisikustamise tõenäosust.

Üldistamine võib olla näiteks täpsuse vähendamine, kategoriseerimine või agregeerimine. Näiteks on või-

malik vähendada tunnuse väärtuste detailsust (täpse aadressi asemel kasutatakse kohaliku omavalitsuse täpsust), täpsed väärtused asendada väärtusvahemike või intervallidega (täpse vanuse asemel kasutada vanusegruppe) ning väärtused grupeerida (täpse ametinimetuse asemel kasutada ametigruppe nagu näiteks tippspetsialistid, tehnikud ja keskastme spetsialistid, teenindus- ja müügitöötajad, juhid). Ka pseudonüümimise protsessis kasutatakse tihti anonüümimise meetodeid, näiteks tunnuste üldistamist või täpsuse vähendamist.

Pärast müra lisamist või üldistamist on mõistlik kasutada $k$-anonüümimist. $k$-anonüümsuse definitsioon ilmus teaduskirjanduses esimest korda 2007. aastal [32], ning see tähendab, et iga kvaasi-identifikaatorite kombinatsioonile vastab anonüümitud andmestikus vähemalt $k$ isikut. Mida suurem on $k$, seda paremini on isikute privaatsus kaitstud, aga seda vähem andmeid jõuab anonüümitud andmebaasi, sest kõik unikaalsed kirjed eemaldatakse. $k$-anonüümsuse saavutamiseks, esmalt tuvastatakse kvaasi-identifikaatorid ning nende põhjal grupeeritakse iga isik vähemalt $k-1$ teise isikuga. Kahjuks kasvab selle probleemi keerukus ning ajamahukus kvaasi-identifikaatorite arvu suhtes.

Kuna ka $k$-anonüümsus ei vähenda alati piisavalt taasidentifitseerimise riski, lisatakse sellele meetodile veel $l$-hajutus ja $t$-lähedus. $l$-hajutus laiendab $k$-anonüümsust, lisades sinna kontrolli, et igas grupis oleks igal tunnusel vähemalt $l$ erinevat väärtust. Kahjuks see meetod ei tööta hästi, kui tunnused grupis on ebaühtlase jaotusega (mõnda väärtust on palju ja teisi vähem kui $l$) või on väheste väärtuste arvuga (näiteks sugu). $t$-lähedus laiendab $l$-hajutust, nõudes et igas grupis oleks vähemalt $l$ erinevat väärtust ja iga väärtus peab esinema nii mitu korda, et iga tunnuse jaotus vastaks jaotusele algses andmestikus.

Juhtumianalüüsina on heaks näiteks LEOSS projekt[25] (Lean European Open Survey on SARS-CoV-2 Infected Patients), mis on üle 10 000 patsiendi anonüümitud avaandmetega register COVID-19 seotud epidemioloogiliste ja kliiniliste uuringute toetamiseks.[26] Anonüümimise läbiviimiseks kasutati anonüümimistööriista ARX[27].

Enne andmete anonüümimist hinnati nende avaldamisega seotud riske ja viidi läbi iteratiivne testanonüümimine sünteetiliste andmete peal. Selle käigus hinnati, kui suur on iga vaadeldava tunnusega seostatud tagasituvastamise risk, ning sõnastati anonüümitud andmetele järgmised nõuded: suurima riskiga atribuutide juures peab rakendama $k$-anonüümimist ($k = 11$), eemaldada tuleb suure $t$-lähedusega ($t < 0.5$) tundlikud tunnused/atribuudid ning erandlikud väärtused. Lisaks tuli veenduda, et vastavad nõuded kehtivad andmete pideval lisamisel [33].

**Turvagarantiid ja jääkriskid.** Põhilisteks rünneteks anonüümitud andmestikele on välja otsimine, seostamine ja järeldamine. Välja otsimine on ühe isiku osade või kõikide väärtuste tuvastamine andmestikust ning on seetõttu kõige suurema privaatsusriive riskiga. Seostamine on ründaja võime seostada vähemalt kaks sama isiku või grupi kirjet samas andmestikus või erinevates andmestikes. Järeldamine on teiste tunnuste põhjal tunnuse väärtuse äraarvamine piisavalt suure tõenäosusega. Kaks viimast rünnet ei tuvasta otseselt andmebaasis üksikisikut, aga nende abil on isiku kohta võimalik midagi teada saada.

Juhuslikkuse lisamise puhul on välja otsimine võimalik, aga tulemused on vähem usaldusväärsed (on tõenäosus, et tunnuse väärtust on teatud ulatuses muudetud). Seostamine on ka võimalik, aga on võimalus, et päris kirje seostatakse ekslikult müraga. Järeldamine võib ka olla võimalik, aga õnnestumistõenäosus on madalam.

Korrektselt saavutatud $k$-anonüümsuse puhul ei tohiks olla võimalik isikut andmestikust välja otsida. Seostamine on võimalik, aga piiratud võimekusega. Järeldusrünne on võimalik, näiteks juhul kui kõik $k$ isikut on samas grupis ja ründaja teab, millisesse gruppi isik kuulub (näiteks kui grupis, kuhu isik kuulub, on kõikide inimeste palgad vahemikus $2500$-$3000$ eurot). $l$-hajutus tegeleb sellega, et deterministlik järeldusrünne ei oleks võimalik, nõudes, et igal tunnusel esineks vähemalt $l$ erinevat väärtust. See muudab järeldusründe tõenäosuslikuks.

Ebapiisavalt anonüümitud avaandmete taasisikustamist on näidatud mitmel pool maailmas [34, 35], eriti kui ründajal on võimalik andmestikku piiramatult ühendada teiste olemasolevate andmestikega, mis taasidentifitseerimist lihtsustavad.

[25] https://leoss.net (viimati külastatud 02.03.2023).
[26] https://leoss.net/data (viimati külastatud 02.03.2023).
[27] ARX – Data Anonymization Tool https://arx.deidentifier.org (viimati külastatud 02.03.2023).

**Juhised rakendajale.** Väga oluline on, et anonüümija tunneks andmestikku hästi. Siis on võimalik valida korralik komplekt kvaasi-identifikaatoreid. On olemas tööriistu, mis selle valikuga toetavad, aga kuna see on väga töömahukas, siis hea andmestiku tundmine tõstab edu tõenäosust.

Juhuslikkuse lisamist ja üldistamist on mõistlik kasutada ainult lisameetoditena (näiteks, et andmestikku $k$-anonüümimiseks ette valmistada). $k$-anonüümimise puhul on oluline, et parameeter $k$ oleks hästi valitud. Kuna tulemist kustutatakse kõik kirjed, mida ei ole piisavalt palju, et moodustada $k$-elemendilist gruppi, siis kui $k$ on liiga suur, võib tulem algset andmestikku halvasti esindada ning analüüsist võivad kaduda olulised andmed. Kui suure $k$ puhul vähendada kvaasi-identifikaatorite arvu, on võimalik $k$-elemendilisi gruppe lihtsamalt moodustada, aga sellisel juhul risk isikute välja otsimisele suureneb.

Kui $k$ on liiga väike, on iga isik grupis statistiliselt liiga oluline ning see võib mõjutada andmeanalüüsi tulemusi. Lisaks, kuna väikeses grupis on tunnustel vähem erinevaid väärtuseid, on lihtsam teha järeldusründeid.

$k$-anonüümimine ei tööta suuremahuliste andmestike peal (näiteks ostusoovituste ja ostude andmestikel).

**Õiguslikud aspektid**

Anonüümimine on selline meetod, mille kohaselt füüsilise isiku tuvastamine ei ole võimalik. Kui siiski füüsilise isiku tuvastamine osutub mingil moel võimalikuks, ei ole tegemist anonüümsete vaid pseudonüümitud andmetega, mille puhul rakendatakse isikuandmete kaitse nõudeid. IKÜMi pp 26 kohaselt ei tuleks andmekaitse põhimõtteid kohaldada esiteks, anonüümse teabe suhtes, nimelt teave, mis ei ole seotud tuvastatud või tuvastatava füüsilise isikuga, ja teiseks, isikuandmete suhtes, mis on muudetud anonüümseks sellisel viisil, et andmesubjekti ei ole võimalik tuvastada või ei ole enam võimalik tuvastada. IKÜMis ei käsitleta sellise anonüümse teabe töötlemist, sh statistilisel või uuringute eesmärgil. (IKÜM, pp 26)

Kõikide kaitsetehnoloogiate, sh ka anonüümimismeetodite kasutamisel tuleb arvestada, et meetodite kaitsetase võib erineda. Samuti võib erineda ka sama meetodi kasutamine erinevates kontekstides, sõltudes näiteks teistest kaitsemeetmetest ja nende tasemetest.[28] Privaatsuskaitse tehnoloogiate meetodite õige rakendamise tagamine on võtmeküsimus vastamaks IKÜMi nõuetele. Käesoleval ajal puudub aga süstemaatiline kirjanduse ülevaade PET-ide kohta, mis keskenduks juriidiliste ja tehniliste nõuete anonüümsuse taseme hindamisel [36].

[28] Vt nt anonüümimismeetodite kasutamise problemaatikaid, Bogdanov, D., Siil, T., Infotehnoloogilised võimalused põhiõiguste kaitsel [2], lk 476.

## Anonüümimine

**ANDMED**

**Inglise keeles:** anonymisation

**Lühidalt:** Anonüümimise käigus eemaldatakse, muudetakse ning teisendatakse andmestikku nii, et tulemist oleks võimalikult raske või võimatu tuletada isikustatud andmeid.

**Arenduse keerukus:** keskmine

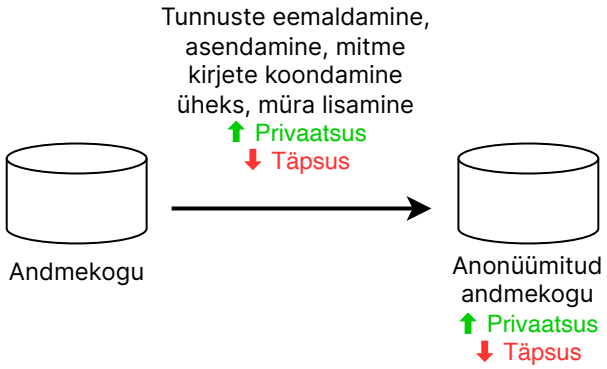
**Ülalpidamise keerukus:** madal

**Täpsus:** ebatäpne (lisatakse müra, eemaldatakse andmeid)

**Privaatsusgarantii:** tõestatavat privaatsusgarantiid ei ole

**Tehnoloogia küpsus:** keskmine

**Ülevaatlik mudel:**




**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: anonüümimise teostajal on väga hea ettekujutas andmestiku omadustest.
2. Turvaeeldus: tuvastatud on otsesed ja kaudsed, kvaasi-identifikaatorid.
3. Turvaeeldus: iga tunnuse jaoks on vastavalt andmetüübile ja jaotusele valitud sobiv tehniline lähenemine.
4. Turvaeeldus: anonüümitud andmestiku töötleja tegevus on piiratud.
5. Jääkrisk: anonüümimine võib muutuda võimalikuks, kui ründaja saab andmebaasi ühendada lisanduvate andmetega

**Rakendusvõimalused:**

1. Avaandmete avaldamine
2. Andmete avaldamine testimiseks või teadustööks
3. Andmete kasutamine poliitika- või üldistes analüüsides
4. Andmete kaasamine protsessidesse nagu klienditugi ja vigade tuvastus, kus välditakse isikute tuvastamist.

**Õiguspraktika:**

1. Anonüümimine kui tehnoloogia ei tarvitse alati tagada väljundandmestiku anonüümsust IKÜM põhjenduspunkt 26 mõistes. Anonüümsuse saavutamiseks tuleb tagada, et isikustatud andmete tuletamine anonüümitud andmestikust ei ole mõistliku pingutusega võimalik (arvestades võimaliku ründaja vahendeid).

**Tuntumad rakendused:**

1. EITaF ja Saksamaa Nakkushaiguste liidu avaldatud LEOSS andmestik Euroopa COVID-19 patsientide kohta

# LEOSS – Euroopa COVID-19 patsientide andmete jagamine anonüümitult

**Lühidalt:** LEOSS on projekt COVID-19 patsientide avaandmete teadlastele kättesaadavaks tegemiseks

**Teostamise aasta:** 2020

**Riik:** Saksamaa

**Omanik:** Haiglad ja muud terviseasutused

**Teostaja:** Koostööprojekt, mille algatajaks olid EITaF (Emerging Infections Task Force) ja Saksamaa nakkushaiguste selts (German Society for Infectious Diseases)

**Süsteemi küpsus:** püsiv juurutus

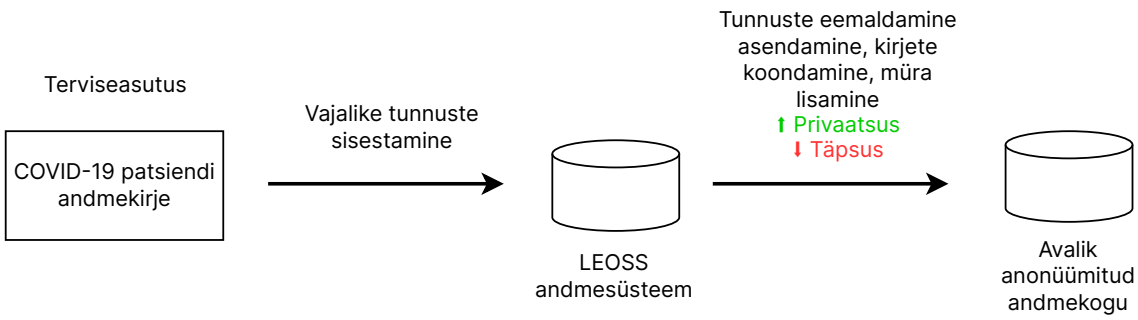
**Privaatsuskaitse tehnoloogiad:**

1. anonüümimine

**Sobivad kasutusjuhtumid:**

avaandmed

**Ülevaatlik mudel:**




**Märkimisväärsed omadused:**

1. Andmete anonüümimisele ja avaldamisele eelnes põhjalik riskianalüüs, testanonüümimine sünteesitud andmetel.
2. Rakendati mitut tehnikat – $k$-anonüümimist ja $t$-lähedusega.
3. Registris on 10 riigist ja enam kui 130 instituudist kogutud patsientide andmed. Anonüümitud andmebaasis on üle 10 000 kirje.
4. Auhinna Open Data Impact Award kandidaat.

### 4.2.3 Piirangutega päringuliidesed

**Lihtsustatult.** Piirangutega päringuliidesed lubavad andmebaasi kohta küsida vaid eelnevalt kokku lepitud küsimusi. Nii kahaneb oht, et küsimuse vastuses on isikustatavaid andmeid.

**Ülevaade ja rakendamine.** Olgu meil andmekogu, millel soovivad päringuid teha mitmeid välised isikud. Kui me lepime kokku päringute tüüpides, mis rahuldavad analüütikute vajadusi ning piiravad üksikkirjete andmist ja lekkimist, siis saame luua masinliidese, mille kaudu selliseid kokkulepitud ja piiratud päringuid esitada saab. Näiteks võime kujutleda infosüsteemi, millele saab edastada päringuna inimese eesnime ning vastuseks antakse kõigi selle nimega isikute vanuste keskmine või siis teade, et sellise nimega isikuid andmekogus ei ole. Liidese kasutajal ei ole võimalik esitada teistsuguseid päringuid vaid ainult muuta lubatud parameetreid vastavalt lubatud vahemikele. Selline lahendus piirab töötlemist, välistades terved kategooriad võimalikke riskantseid päringuid. Sellised piirangud on erinevad infoturbe meetmetega seatud juurdepääsu piirangutest, mis välistavad süsteemi kasutamise näiteks volitamata isikutele.

Päringute paindlikkuse määrab andmekogu omanik. Kõige piiratumal juhul on päringud ettemääratud ning parameetriteta. Sellisel juhul sõltub päringu vastus vaid andmekogu sisust. Keerulisemal juhul saab kasutaja päringule ette anda lisa valikuid, mis seda täpsustavad.

Sinna vahele jäävad mitmed vahepealsed variandid, näiteks:

1. rakendusliides, millele saab anda parameetreid (nt lähteandmete filtreerimiseks);
2. päringukeel, milles on lubatud alamosa (nt SQL ilma filtreerivate *WHERE*-lauseteta);
3. liides, mis keeldub andmast vastuseid, mis on arvutatud alla kolme andmekogu kirje pealt.

Piiratud päringuliideste kasutamine on maailmas populaarne andmetele juurdepääsu loomise viis, seda eriti just avaandmete ja statistika puhul. Eesti kõige olulisem piiratud päringukeskkonna rakendus on Statistikaameti statistika andmebaas[29]. Seal avaldatakse piiratud päringukeskkonna abil koondtulemusi Eesti statistikast. Isikuandmete kaitse tagatakse sellega, et päringute vastustes on vaid koondtulemid (kasutatud on andmete agregeerimist).

Avaandmete teabevärav[30] võimaldab jagada andmeid taotluste põhiselt, kus päringu tegija on tuvastatud. Lisaks on teabevärava rakendusliidese kasutamine võimalik vaid TARA kaudu autenditud kasutajatel, seega on võimalik vältida rakendusliidese kuritarvitamist ja omada ülevaadet andmete kasutajatest.

**Turvagarantiid ja jääkriskid.** Olgu meil andmebaas, milles on isikustatavad andmeid ja millele on üles seatud piiratud päringuliidese tehnoloogiaga kasutaja- või masinliides. Vaid selle tehnoloogiaga ei ole võimalik anda kindlat turvagarantiid, et mitme päringu tulemused on omavahel kombineerides anonüümsed ja ei ole isikustatavad.

Peamine jääkrisk on, et osav analüütik esitab mitu päringut ja neid kokku pannes suudab saada mõne isikustatud tulemuse. Mida paindlikum on liides, seda suurem on sellise andmete lekke oht. Kui kasutaja päringute arv ei ole piiratud, siis saab ta tõenäoliselt üle mitmete päringute kätte piisavalt infot, et andmekogu mingis ulatuses taastada. Selliste rünnete läbiviimiseks on olemas ka poolautomaatsed tööriistad [37], seega nende tõenäosus ajas kasvab.

**Juhised rakendajale.** Selliste päringuliideste ehitamiseks on võimalik teha projekt, mis kaardistab andmekogu koosseisu, ärivajadused ning selle põhjal hindab, milliseid päringuid lubatakse. Kui on võimalik fikseerida päringud ja põhjendada, et lubatud päringute kõikvõimalikud tulemused ei saa isikustatavaid andmeid väljastada, siis võib infosüsteemi ehitada tavapärasel viisil.

Kui analüüsi käigus leitakse, et teatud tingimustel võivad päringu tulemused olla isikustatavad, tuleb kas päringute paindlikkust vähendada, või lisada mõni täiendav privaatsuskaitse tehnoloogia, arvestades vajadusega säilitada süsteemi kasutatavus.

Andmete koondamise (peatükk 4.2.2) rakendamisel saaks liidese kaudu küsida vaid koondtulemusi. Sellisel juhul on tähtis, et kasutaja ei saaks filtrite abil küsida koondtulemust vaid ühest kirjest, sest see avaldaks lähteandmebaasi kirje. Näiteks ei tohiks saada pärida kõrgeima palgaga inimese keskmist sis-

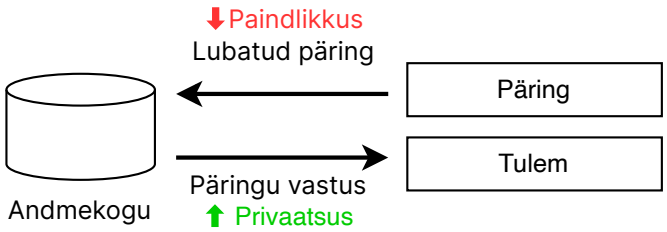
[29] Statistikaameti statistika andmebaas. https://andmed.stat.ee/et/stat (viimati külastatud 03.03.2023)
[30] Avaandmete teabevärav, https://avaandmed.eesti.ee (viimati külastatud 03.03.2023)

setulekut. Müra lisamisega (vt peatükid 4.2.2 ja 4.2.5) saavutataks see, et päringute tulemused üle mitme päringu ei oleks enam omavahel seostatavad ning seega ei oleks selline rünne enam nii lihtsalt teostatav.

Privaatsuseelarve järgimine tähendaks, et iga kasutaja saaks andmekogu kohta esitada vaid piiratud arv päringuid. Mida täpsemalt see arv on seotud andmekogu kirjete struktuuri ning päringute iseloomuga, seda parem on kaitse. Efektiivseid privaatsuseelarve meetodeid saab samuti ehitada diferentsiaalprivaatsuse meetodeid (vt peatükk 4.2.5) kasutades.

| **Piirangutega päringuliidesed** | **ANDMED** |
|---|---|
| **Inglise keeles:** restricted query interfaces | |
| **Lühidalt:** Piirangutega päringuliidesed lubavad andmebaasi kohta küsida vaid eelnevalt kokku lepitud küsimusi. Nii kahaneb oht, et küsimuse vastuses on isikustatavaid andmeid. | **Arenduse keerukus:** madal |
| | **Ülalpidamise keerukus:** madal |
| | **Täpsus:** täpne |
| | **Privaatsusgarantii:** tõestatavat privaatsusgarantiid ei ole |
| | **Tehnoloogia küpsus:** kõrge |

**Ülevaatlik mudel:**




**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: Päringuliides peab olema võimalikult paindumatu.
2. Turvaeeldus: Päringute hulka tuleb piirata ning seirata.
3. Jääkrisk: paindliku päringuliides puhul tekib oht, et paljude kavalate päringute abil saab kasutaja isikustatud tulemusi.

**Rakendusvõimalused:**

1. Avaandmete teenused (ptk 6.3)
2. Andmete piiratud jagamine näiteks teadustööks või mitte-ärilistel eesmärkidel
3. X-tee turvaserveri teenused

**Õiguspraktika:**

1. Kehtivad tavapärased andmekaitsenõuded.
2. Kui päringu esitajal on võimalik saada isikustatud teavet, peab tal olema selleks vastav õiguslik alus.

**Tuntumad rakendused:**

1. Statistikaameti statistika andmebaas (piiratud päringuliides on juurutatud koos andmete koondamisega (ptk 4.2.2)
2. Avaandmete teabevärav

### 4.2.4 Analüütiku töökohad

**Lihtsustatult.** Analüütiku töökohal saab kasutaja andmeid vabalt töödelda. Tema tegevust jälgitakse ja analüüsi tulemid kontrollitakse enne kasutajale üle andmist, et nendes poleks isikustatavaid andmeid.

**Ülevaade ja rakendamine.** Esineb olukordi, kui päringuid ei ole võimalik ette kokku leppida (nt on tegemist teadustööga või algoritmide arendusega). Selles olukorras saab andmekogu haldaja üles seada analüütiku töökoha, mida analüütik saab kasutada kontrollitud keskkonnas kaug- või füüsilise pääsuga. Andmekogust tehakse analüütiku töökohta ühendus või väljavõte.

Analüütik saab kasutada tavapäraseid andmeteaduse tööriistu või ka omaenda algoritme ja lahendusi. Küll aga ei saa kasutaja analüüsi tulemusi viia sellest arvutist välja enne kui automaatne süsteem või inimene need üle vaatab.

Nii kaug- kui füüsilise ligipääsu juures saab kasutaja panna oma andmetöötluse tulemused ülevaatuseks eraldi keskkonda ning pärast ülevaatust edastatakse need kasutajale kas digitaalselt või andmekandjal. Ülevaatuse käigus veendub analüütiku töökoha teenuse osutaja, et välja viidavad andmed vastavad teenuse tingimustele ja poliitikatele.

Eestis on analüütiku töökoht juurutatud Statistikaametis teadlase töökoha teenusena[31]. Statistikaamet on arendamas sellest ka täiuslikumat versiooni. Tartu Ülikooli Genoomika instituut ehitab analüütiku töökohta Eesti Geenivaramu andmete kasutajale.

**Turvagarantiid ja jääkriskid.** Teenuse kasutajal on vaba juurdepääs andmekogu andmetele ning vabadus neid töödelda. See inimene näeb kõike, mida talle on töötlemiseks valmis seatud. Kui need andmebaasid on isikustatavad, siis on kindlate turvagarantiide tagamine võimatu.

Kaks peamist ohtu on

a) nõrk ülevaateprotsess, mis ei näe, et väljastatavad andmed võivad olla isikustavad ja
b) andmete väljaviimine teisi kanaleid pidi, näiteks ekraanil olevatest andmetest foto- või videosalvestust tehes).

Mõlemat kanalit pidi on võimalik andmekogu andmeid välja viia ning see on võimekale andmeanalüütikule pigem lihtne töö. Suuremate andmete puhul on ekraanipiltide abil see küll ebamugav, kuid ka üks nime ja kompromiteeriva fakti seos võib olla liialt suur leke.

**Juhised rakendajale.** Kõige olulisem samm on valida, millised andmestikud analüütiku töökoha kaudu kättesaadavaks teha. Kui need on isikustatavad andmed või muul moel konfidentsiaalsed, tuleb rakendada täiendavaid privaatsuskaitse tehnoloogiaid. Sobivad näiteks anonüümimine (ptk 4.2.2, sünteetiliste andmete genereerimine (ptk 4.2.7 ja diferentsiaalprivaatsus (ptk 4.2.5. Analüütiku töökoha tegevuste pidev logimine ei hoia väärkasutust ära, kuid võib olla seda pärssiva toimega.

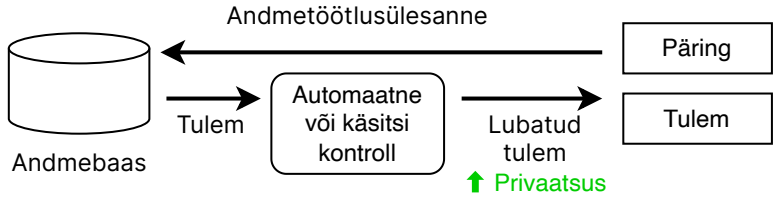
Organisatoorsetest meetmetest sobivad lepingulised piirangud ja veenvad trahvid teenuse tingimuste rikkumise eest. Füüsilise pääsuga süsteemi puhul saab rakendada turvakaameraid, millel võib olla ennetav toime. Kaugpääsuga süsteemi puhul seda võimalust küll ei ole.

Platvorme analüütiku töökohtade jaoks on tootestatud ning turul on sellele mitmeid pakkujaid. Analüütiku töökoht on võimalik ehitada ka laiatarbe tehnoloogiaid osavalt kombineerides.

[31] Konfidentsiaalsete andmete kasutamine teaduslikul eesmärgil. https://www.stat.ee/et/avasta-statistikat/kusi-statistikat/konfidentsiaalsete-andmete-kasutamine-teaduslikul-eesmargil (viimati külastatud 17.01.2022)

| **Analüütiku töökohad** | **ANDMED** |
|---|---|
| **Inglise keeles:** analyst sandboxes | |
| **Lühidalt:** Analüütiku töökohal saab kasutaja andmeid vabalt töödelda. Tema tegevust jälgitakse ja analüüsi tulemid kontrollitakse enne kasutajale üle andmist, et nendes poleks isikustatavaid andmeid. | **Arenduse keerukus:** madal<br>**Ülalpidamise keerukus:** madal<br>**Täpsus:** täpne<br>**Privaatsusgarantii:** tõestatavat privaatsusgarantiid ei ole<br>**Tehnoloogia küpsus:** keskmine |

**Ülevaatlik mudel:**




| | |
|---|---|
| **Turvaeeldused ja jääkriskid:**<br>1. Turvaeeldus: Kanalid andmete töökohast välja viimine peab olema võimalikult piiratud.<br>2. Turvaeeldus: Andmete väljaviimise kontroll peab võimalikult hästi tuvastama isikustatavaid andmeid.<br>3. Jääkrisk: siiski näeb analüütik töökeskkonnas kõiki andmeid ning saab sealt fakte meelde jätta või pildistada. | **Rakendusvõimalused:**<br>1. Andmebaasidelele juurdepääs teadus- või õppetööks |
| **Õiguspraktika:**<br>1. Isikuandmete töötlemise puhul tuleb järgida privaatsus- ja andmekaitse nõudeid. Näiteks peab analüütiku töökoha kasutajal olema õiguslik alus andmete laialdaseks töötlemiseks. | **Tuntumad rakendused:**<br>1. Statistikaameti analüütiku töökoht<br>2. Eesti Geenivaramu Andmepuur |

### 4.2.5 Diferentsiaalprivaatsus

**Lihtsustatult.** Diferentsiaalprivaatsus teeb päringu vastused juhuslikuks nii, et küsija ei saa aru, milliste isikute andmete pealt päring tehti.

**Tehnoloogia ülevaade ja rakendamine.** Diferentsiaalprivaatsus on andmebaasipäringutele vastamise mehhanismi (näiteks päringusüsteemi) omadus oma väljundis avaldada infot üksikute kirjete kohta ainult vähesel määral. Omadust infot mitte avaldada või avaldada ainult vähesel määral on ilmselt võimalik defineerida mitmel erineval, „intuitsioonile vastaval" viisil. Diferentsiaalprivaatsuse omadus kombineerib intuitsiooni esiteks realiseeritavusega ja teiseks komponeeritavusega. Komponeeritavus on oluline suurte süsteemide analüüsil – kui me soovime süsteemi osade diferentsiaalprivaatsusest matemaatiliselt korrektselt järeldada midagi terve süsteemi diferentsiaalprivaatsuse kohta.

*Diferentsiaalprivaatsuse omadus* võib mingil andmebaasipäringutele vastamise mehhanismil olla või mitte olla. Diferentsiaalprivaatseid mehhanisme on edukalt kasutatud agregeerimistulemuste teatavaks tegemiseks kas avalikkusele või mingile konkreetsele infotarbijale juhul, kui allolev andmebaas on *suur*.

Üks tuntumaid näiteid on USA 2020. aasta rahvaloenduse tulemuste avaldamine [38]. Samuti kasutavad Apple [39] ja Google [40] diferentsiaalprivaatseid meetodeid, mille abil nad saavad teada suurte kasutajarühmade agregeeritud eelistusi, ilma üksikute kasutajate kohta täpseid andmeid kogumata.

Diferentsiaalprivaatsuse definitsioon on kvantitatiivne, ta defineerib andmete seostamatuse (näiteks isikuga) sõltuvalt mittenegatiivsest parameetrist $\varepsilon$. Siin $\varepsilon = 0$ tähendab, et päringuvastus ei sõltu andmebaasist, st privaatsus on absoluutne. Kui $\varepsilon$ kasvab, siis privaatsus väheneb; kui $\varepsilon \to \infty$, siis muutuvad nõuded andmebaasipäringutele vastamise mehhanismile olematuks.

*Komponeeritavus* tähendab, et kui teeme kaks andmebaasipäringut, millest ühele vastatakse mehhanismiga, mille diferentsiaalprivaatsus on $\varepsilon_1$, ja teisele mehhanismiga, mille diferentsiaalprivaatsus on $\varepsilon_2$, siis mõlema päringuvastuse diferentsiaalprivaatsus on kokku $\varepsilon_1 + \varepsilon_2$ (mainime, et liitmine ei ole ainus viis, kuidas diferentsiaalprivaatsuse määrad kombineeruda võivad).

**Täpsusest.** Diferentsiaalprivaatsed saavad olla ainult mehhanismid, mis lisavad müra (va juhul, kui päringuvastus üldse andmebaasist ei sõltu). Müra lisamine vähendab päringuvastuse täpsust. Parameeter $\varepsilon$ ei iseloomusta, kui palju väheneb täpsus; ta iseloomustab ainult privaatsust.

Müra on võimalik lisada andmebaasile enne päringuvastuse väljaarvutamist, päringuvastusele enne selle tagastamist, või mingitele sobivatele vahetulemustele päringuvastuse väljaarvutamise käigus. Sel viisil võime saada mehhanisme, mille privaatsusparameeter $\varepsilon$ on üks ja seesama, aga täpsus (või kasulikkus; see sõltub väga palju sellest, mida päringuvastusega edasi tehakse) on väga erinev.

Suur osa diferentsiaalprivaatsuse-alasest teadustööst püüab leida teatud päringuklasside jaoks mehhanisme, mille privaatsuse-täpsuse suhe oleks võimalikult hea [41].

**Ajaloost.** Diferentsiaalprivaatsuse definitsiooni pakkusid 2006. aastal välja Dwork, McSherry, Nissim ja Smith [42], kes selle eest 2017. aastal Gödeli preemia said. Definitsioonis kajastuvad varasemad tulemused statistiliste andmebaaside turvalisuse vallast, muuhulgas tähelepanek, et andmebaasipäringute vastused, mis on antud kas ilma mürata või väga väikese müraga lubavad ründajal, kes saab esitada piisavalt palju päringuid (mis on juhuslikku laadi), enda jaoks taastada suure osa andmebaasist [43]. Samuti võib keeruline olla filtreerida päringuid selliselt, et ründaja mingit teatud tundlikku väärtust teada ei saaks [44].

**Turvagarantiid ja jääkriskid.** Kõige tähtsam turvaeeldus on õige $\varepsilon$ parameetri valik. Kui see on konkreetselt tehtud, siis on andmete isikutega seostamatuse matemaatiliselt tõestatav. Diferentsiaalprivaatsus väga võimas tööriist ning üks väga vähestest mittekrüptograafilistest privaatsuskaitse tehnoloogiatest, millel on niivõrd tugev lubadus.

Paraku ei ole $\varepsilon$ interpreteerimine mingites „intuitiivsemates" terminites ja seega ka valik triviaalne. See nõuab andmestiku ning päringumehhanismi head tundmist ja selleks on soovitatav kasutada spetsialistide ja tööriistade abi. Leidub viise $\varepsilon$ sidumiseks mingite atribuutide äraarvamise tõenäosuse või täpsusega [45].

**Juhised rakendajale.** Esimene samm on andmestiku ja tunnuste valik. Kui diferentsiaalprivaatsed meh-

hanismid mingis rakenduses kasutusele võtta, seda siis ilmselt mingite agregeerimistulemuste avaldamiseks. Kõigepealt tuleb otsustada, millised väärtusi andmebaasis me kui palju kaitsta soovime, ja väljendada seda mingi väärtusena $\varepsilon$. Kui me soovime kaitsta eri tüüpi väärtusi, siis võib meil veel olla tarvis otsustada, kui palju rohkem me neist ühtesid võrreldes teistega kaitsta soovime [46].

Google on avaldanud teegid diferentsiaalprivaatsete päringumehhanismide realisatsioonidega teatud agregatsioonioperatsioonide jaoks[32]. Keerulisemate operatsioonide jaoks leidub avalikustatud akadeemilisi teeke [47][33].

Teine samm on sobiva täpsusega diferentsiaalsuse mehhanismi ja parameetri $\varepsilon$ valik. Kui me oleme sobivad arvulised väärtused ja kompromissid välja valinud, siis tuleb saadavalolevate seast leida päringusüsteemile või analüüsialgoritmile sobiv diferentsiaalprivaatselt vastamise mehhanism. Tuleb valida sobivat seostamatuse taset pakkuv $\varepsilon$.

Seejärel tuleb mehhanismi testida ja otsustada, kas seostamatus ja täpsus, mille sealt saame, on piisav. Võib-olla piisab siin üldisest mehhanismist, mis võtab ilma privaatsuseta mehhanismi ja lisab selle tulemusele müra määral, mis sõltub $\varepsilon$-st ja päringu kujust. Kui sellest aga ei piisa, tuleb otsida parema täpsusega mehhanism.

Kolmas samm on korrigeerimine. Juhime tähelepanu ka sellele, et diferentsiaalprivaatsus on mõeldud kaitsma andmebaase, kus kirjed on üksteisest sõltumatud. Kui kirjeid ei saa üksteisest sõltumatuteks lugeda (näiteks on tegemist inimeste geeniinfoga, aga andmebaasis olevad inimesed on omavahel sugulased), siis on privaatsusgarantii väiksem kui $\varepsilon$ väärtusest arvata võiks. Rusikareegel on, et kui kirjed on üksteisest sõltuvad nii umbes rühmadena suurusega $k$, siis on efektiivne privaatsusgarantii umbes $k \cdot \varepsilon$.

Lisakaalutlusena – müra võib lisada andmebaasile enne päringuvastuse väljaarvutamist või päringuvastusele peale selle väljaarvutamist. Esimesel juhul võib olla võimalik süsteem üles seda nii, et päringuvastuse väljaarvutaja ja andmebaasipidaja ei näegi ilma mürata andmebaasi

Teisel juhul peab päringuvastuse väljaarvutajal olema piiranguteta ligipääs andmebaasile, s.t. arvutaja on usaldatud osapool. Usaldusenõuete vähendamiseks on võimalik lisaks kasutada turvalise ühisarvutuse (ptk 4.2.10) või usaldatavate täitmisekeskkondade (ptk 4.2.8) tehnikaid.

Kui andmebaas sisaldab informatsiooni paljude isikute kohta, kusjuures isikud ise annavad need andmed, siis võivad isikud lisada müra juba enne seda, kui nad oma andmed edastavad. Sellist tehnikat kutsutakse näiteks lokaalseks diferentsiaalprivaatsuseks (*local differential privacy*). Teine näide on sotsiaalteadustest tuntud, piinlikele küsimustele ausamate vastuste saamiseks kasutatav juhuslikustatud vastuste tehnika (*randomized response technique*) on lokaalse diferentsiaalprivaatsuse üks vorm.

**Õiguslikud aspektid.** Diferentsiaalprivaatsus on kujunemas oluliseks andmekaitsetehnikaks [48]. Seega on rohkem uuritud ka selle õiguslikke aspekte [49]. Kirjanduses leidub artikleid, mis seostavad anonüümsust ja isikuandmete kaitse taset diferentsiaalprivaatsuse meetodiga [38]. Avaldatud on väiteid, et diferentsiaalprivaatsusega teostatud masinõpe võiks võiks vastata IKÜMi nõuetele [48].

[32] https://github.com/google/differential-privacy (viimati külastatud 01.03.2023).
[33] https://ektelo.github.io (viimati külastatud 01.03.2023).

## Diferentsiaalprivaatsus — ANDMED

**Inglise keeles:** differential privacy

**Lühidalt:** Diferentsiaalprivaatsus teeb päringu vastused juhuslikuks nii, et küsija ei saa aru, milliste isikute andmete pealt päring tehti.

**Arenduse keerukus:** kõrge

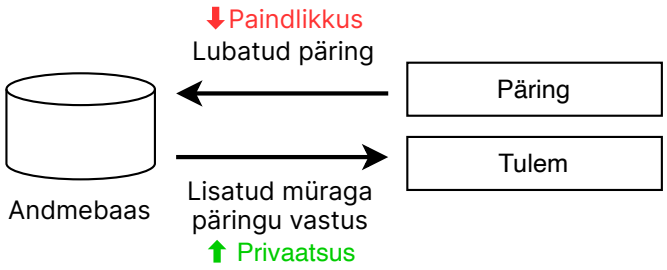
**Ülalpidamise keerukus:** madal

**Täpsus:** ebatäpne (sõltub lisatavast mürast)

**Privaatsusgarantii:** matemaatiliselt tõestatav

**Tehnoloogia küpsus:** keskmine

**Ülevaatlik mudel:**




**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: vaja on valida päringusüsteemi või analüüsialgoritmi jaoks sobiv diferentsiaalprivaatsuse mehhanism.
2. Turvaeeldus: vaja on õigesti valida parameeter $\varepsilon$, mis määrab seostamatuse ja täpsuse taseme.

**Rakendusvõimalused:**

1. Küsitluste läbiviimine.
2. Andmebaaside avaldamine ja kasutamine uuringuteks ja teadus- või õppetööks.
3. Avaandmete teenused (ptk 6.3).

**Õiguspraktika:**

1. Kirjanduses on leitud, et diferentsiaalprivaatne masinõpe võiks olla meetod, mis vastab IKÜMi nõuetele.
2. Kohtupraktika diferentsiaalprivaatsuse osas on vähene või puuduv.

**Tuntumad rakendused:**

1. Diferentsiaalprivaatsed avaandmed Ameerika Ühendriikide rahvaloenduse andmete põhjal (ptk 5.1.4).
2. Apple iOS, macOS, Google ja Microsoft diferentsiaalprivaatsed teenused kasutaja eelistuste õppimiseks.
3. Uber juhtide ja kasutajate asukoha andmete töötlemise eksperiment.

### 4.2.6 Liitõpe

**Lihtsustatult.** Liitõpe on masinõpe, kus iga andmeomanik treenib mudeli oma andmete peal ja siis pannakse mudelid kokku.

**Ülevaade ja rakendamine.** Liitandmebaasiks (*federated database*) kutsutakse mitut autonoomset andmebaasisüsteemi, mis on virtuaalselt seotud üheks andmestikuks. Liitõppe heaks küljeks on see, et andmed ei lahku vastutava töötleja juures, välja saadetakse ainult agregeerimistulemumsed või masinõppe mudel. See meetod sobib väga hästi kokku IKÜM-i andmete minimeerimise põhimõttega.

Andmebaase on võimalik jagada horisontaalselt ja vertikaalselt. Horisontaalselt jagatud baaside puhul on osa kirjeid ühes andmebaasis ja osa kirjeid teises. Tavaliselt on sellisel juhul andmemudelid sarnased. Näiteks kui ettevõte hoiab oma erinevates riikides asuvate esinduste töötajate andmeid vastavas riigis asuvas baasis, on tegemist horisontaalselt jagatud baasiga, sest igas baasis sisalduvad samad tunnused erinevate inimeste kohta. Teiseks horisontaalselt jagatud baasi näiteks on erinevate riikide geenipangad, mis sisaldavad enam-vähem samu tunnuseid oma doonorite kohta, aga igas baasis on erinevate isikute andmed.

Vertikaalselt jagatud baaside puhul on sama isiku erinevad andmed jagatud erinevatesse baasidesse. Sellisel juhul on olemas mingi tunnus, mille abil on võimalik andmeid erinevates baasides linkida. Näiteks on siin riigi kogutavad andmed, mida hoitakse erinevates riigiasutustes. Eesti riiklikud andmekogud moodustavad vertikaalselt tükeldatud liitandmebaasi (näiteks haridusandmed on EHISes, terviseandmed perearsti juures ja haiglate andmestikes, raviarve andmed Tervisekassas, sõidukite andmed Transpordiameti andmestikus). Isiku andmete ühendamine toimub isikukoodi abil.

Liitstatistika on liitõppe lihtsam vorm, kus andmeid ei kasutata masinõppe mudeli treenimiseks vaid nende põhjal tehakse lihtsamat statistikat. Võimalikult palju arvutusi tehakse ära autonoomsetes andmebaasides ja tulemused agregeeritakse keskselt, näiteks andmeteadlase juures. Liitstatistika tegemiseks ei ole tingimata vaja liitandmebaase, seda on võimalik kokku panna ka n-ö käsitsi. Liitstatistikat saab kasutada nii vertikaalselt kui horisontaalselt tükeldatud andmetel. Oluline on tähele panna, et liitstatistika puhul ei piisa lihtsalt erinevates tippudes sama arvutuse tegemiseks. Näiteks ei ole võimalik kahest erinevast baasist pärit sama tunnuse väärtuste keskmisest arvutada üldist keskmist. Selle võimaldamiseks saavad baasid saata andmeteadlasele tunnuse väärtuste summa ja arvu. Selle abil saab andmeteadlane arvutada keskmise.

Liitõppe puhul peavad autonoomsed baasid olema seotud üheks baasiks. Tsentraliseeritud süsteemi puhul koordineerib keskne server tööd ja agregeerib tulemused. See server vastutab tippude valiku eest treenimise alguses ning mudeliuuenduste agregeerimise eest protsessi lõpus. Siin on keskne server pudelikaelaks.

Tsentraliseeritud süsteemis valib keskne arvuti treenimiseks mudeli ja saadab selle andmete hoidjatele. Mudel peab olema valitud nii, et seda oleks võimalik teise samasugusega agregeerida. Andmetipud treenivad lokaalselt mudeli ja saadavad tulemuse tagasi. Keskne analüüsisüsteem agregeerib saadud mudelid. See süsteem võib kaasata nii võrdseid andmearvuteid kui servtöötluse seadmed (ka läbisegi).

Detsentraliseeritud süsteemi koordineerivad autonoomsed baasid iseendid ja agregeerivad tulemused. Sellisel juhul kaob keskne nõrk lüli, sest mudeliuuendusi jagavad kõik baasid omavahel. Kahjuks aga on see süsteem keerukam ja võrgutopoloogia võib mõjutada jõudlust, sest kõik süsteemi osapooled ei ole tavaliselt omavahel seotud ning kõik ei pruugi olla ka ühtlaselt hea võrguühendusega.

Liitstatistika ja liitõpe töötavad väga hästi ka siis, kui andmemahud on väga suured ja andmete liigutamine kesksele andmeteadlasele oleks seetõttu raskendatud. Liitõpet on raske (vahel ka võimatu) kasutada andmestiku vertikaalse tükelduse korral.

Google klaviatuuri Gboard kasutab liitõppe abil treenitud masinõppe mudeleid sisestatava prognoosimiseks [50], emodzide ja piltide soovitamiseks [51] ning grammatika kontrollimiseks.

Apple kasutab liitõpet Siri kõnetuvastusroboti arendamiseks [52]. Meta on liitõppe rakendamiseks koostanud raamistiku, kus tulemuste agregeerimiseks saab kasutada usaldatud käivituskeskkondi. Liitõpet kasutavad ka näiteks IBM, NVIDIA, WeBank [53].

**Turvagarantiid ja jääkriskid.** Liitstatistika puhul liigud baasidest andmeteadlaseni veidi rohkem andmeid kui ainult agregeerimistulemus (nii nagu näiteks keskmise puhul). Keerukamate arvutuste ja algoritmide puhul on lisainfo kogus veelgi suurem. Samas ei ole see kaugeltki nii suur, kui siis, kui kesksesse baasi oleks saadetud kõigist tippudest terved andmestikud.

Kuna tavalise liitõppe puhul saadetakse mudel kas kesksse serverisse või teisele tipule, on oht, et mudeli parameetrid lekivad. Nii on võimalik koguda infot isikute kohta, kes on algses andmestikus. On näidatud, et masinõppemudeleid on võimalik pöörata. See tähendab, et näiteks kui ründajal on masinõppemudel ja veidi lisainfot isiku kohta, saab ta edukalt ennustada isiku mingite teiste tunnuste väärtuseid [54].

Tulemuste turvalisemaks agregeerimiseks saab kasutada turvalist ühisarvutust (peatükk 4.2.10) või usaldatud käivituskeskkondi (peatükk 4.2.8) [55], kuid hetkel on nende kasutusjuhud haruldased. Hispaania keele tekstiprognoosi mudeli treenimisel on Google kasutanud liitõppega koos ka diferentsiaalprivaatsust (peatükk 4.2.5) [56].

**Juhised rakendajale.** Üheks enimkasutatud liitõppesüsteemiks on TensorFlow Federated[34]. Selle süsteemi puhul ei pea andmeteadlane ise liitõppe meetodeid implementeerima vaid saab valida olemasolevate liitõppe algoritmide hulgast sobiva valida, et teha kergemaid ülesandeid nagu mudelite treenimine ja kasutamine. Tundma peab TensorFlowd ja programmikood peab olema kujutatav TensorFlow graafina. Kohaliku agregeerimise peab implementeerima andmeteadlane, mudelite liitõppe süsteemiks koondamisega ning omavahelise agregeerimisega tegeleb TensorFlow Federated.

Tehnilise poole pealt on liitõpet üldiselt raske kasutada väga iteratiivsete algoritmide (näiteks stohhastilise gradientlaskumise puhul). Sellisel juhul on vaja väga väikese latentsi ja suure läbilaskevõimega võrguühendusi tippude vahel.

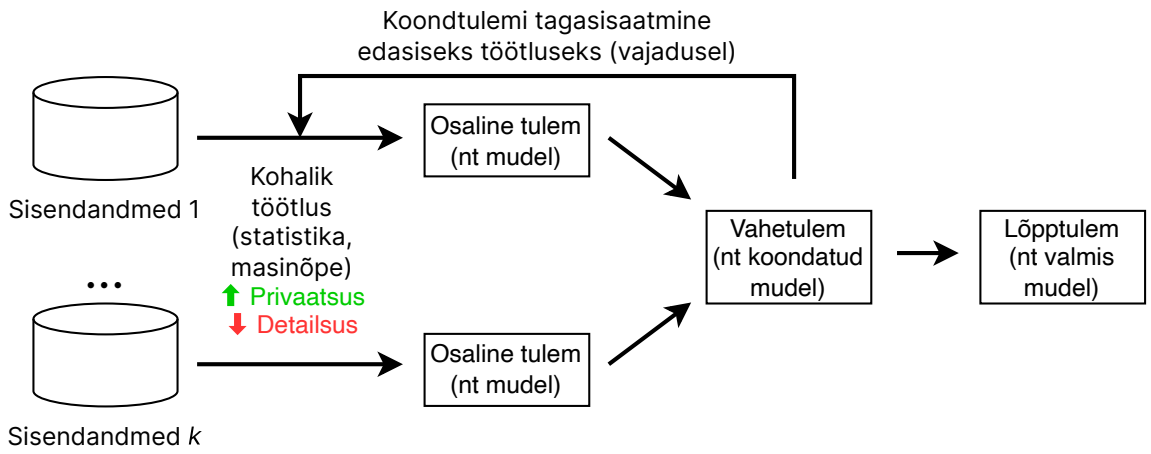
**Õiguslikud aspektid**

Isikuandmete töötlemisel tuleb järgida jurisdiktsioonis kehtivaid privaatsus- ja andmekaitse nõudeid. Olenevalt piirkonnast võivad näiteks liitõppele kohalduda ka tehisintellekti reguleerivate õigusaktide nõuded. Kirjanduses on leitud, et nii nagu mistahes tehisintellekti või masinõppe süsteemi puhul, oleks otstarbekas ka liitõppe puhul koostada privaatsus- ja andmekaitse mõjuhinnang, et võimalikke kaasnevaid riske oleks juba süsteemi arendamisel võimalik parimal võimalikul moel kahandada.

[34] TensorFlow Federated https://www.tensorflow.org/federated/get_started (viimati külastatud 28.02.2023)

| **Liitõpe** | **ANDMED** |
|---|---|
| **Inglise keeles:** federated learning | |
| **Lühidalt:** Liitõpe on masinõpe, kus iga andmeomanik treenib mudeli oma andmete peal ja siis pannakse mudelid kokku. | **Arenduse keerukus:** keskmine |
| | **Ülalpidamise keerukus:** keskmine |
| | **Täpsus:** ebatäpne (mudelite ühendamine on ebatäpsem kui treenimine ühise andmestiku pealt) |
| | **Privaatsusgarantii:** tõestatavat privaatsusgarantiid ei ole |
| | **Tehnoloogia küpsus:** keskmine |

**Ülevaatlik mudel:**




| | |
|---|---|
| **Turvaeeldused ja jääkriskid:**<br>1. Turvaeeldus: Tuleb veenduda, et koondatud ja jagatud osalised mudelid ja statistilised vahetulemid ei lekiks (nt kas isik oli andmestikus või mitte).<br>2. Jääkrisk: vahetulemuste kaudu lekib isikustatud andmeid | **Rakendusvõimalused:**<br>1. Andmeanalüüs horisontaalselt tükeldatud andmebaaside puhul ilma andmeid jagamata. |
| **Õiguspraktika:**<br>1. Märkimisväärseid pretsedente ei õnnestunud leida. | **Tuntumad rakendused:**<br>1. Androidi õppiv klaviatuur Gboard. |

# Androidi õppiv klaviatuur Gboard

**Lühidalt:** Android telefonide klaviatuur Gboard õpib kasutajate harjumuste põhjal nende sisestust ennustama ja ei kogu selleks kõiki andmeid Google'ile kokku.

**Teostamise aasta:** 2017

**Riik:** Ameerika Ühendriigid

**Omanik:** Google

**Teostaja:** Google

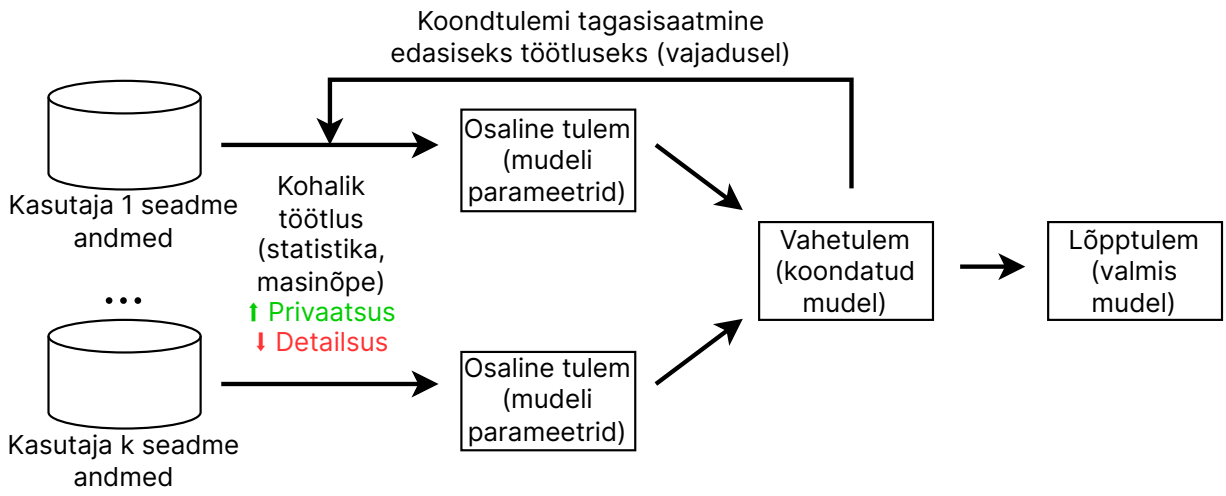
**Süsteemi küpsus:** püsiv juurutus

**Privaatsuskaitse tehnoloogiad:**

1. liitõpe
2. diferentsiaalprivaatsus

**Sobivad kasutusjuhtumid:**
privaatne analüütika

**Ülevaatlik mudel:**




**Märkimisväärsed omadused:**

1. Google Play Store põhjal on rakendust alla laaditud üle viie miljardi korra.
2. Google oli esimesi liitõppe juurutajaid, kuid seda on arendama, mugandama ja kasutama hakanud ka teised suured tehnoloogiaettevõtted.

### 4.2.7 Sünteetiliste andmete genereerimine

**Lihtsustatult.** Sünteetilised andmed on juhuslikud andmed, mis näevad välja nagu pärisandmed.

**Ülevaade ja rakendamine.** Andmesüntees on andmete genereerimine. Sünteetilised andmed võivad olla loodud mingite reeglite, seaduspärasuste või pärisandmete järgi. Selles alajaotises vaatleme pärisandmetele tuginevat andmete genereerimist.

Andmesünteesi lihtsamaid vorme on kasutatud aastakümneid. Andmete imputeerimine on meetod, mille abil statistikas asendatakse puuduvaid väärtusi. Selleks, et hiljem oleks neid andmeid võimalik analüüsis kasutada, on vaja, et need sarnaneksid mingil määral päriselu andmetega. Tavaliselt uuritakse välja, mis on vastava tunnuse jaotus ja valitakse sealt jaotusest juhuslikult andmed. Kahjuks ei pruugi sedasi genereeritud andmed olla vastavuses ülejäänud tunnustega.

Sünteetiliste andmete genereerimisel kasutatakse statistilisi, masinõppe meetodeid ja näiteks ka süvaõpet, et luua erinevate tunnuste põhjal mudel või neurovõrk, mis suudab talletada ka keerukamaid jaotusi ning tunnustevahelisi seoseid. Juhuslikkust lisades luuakse selle mudeli põhjal omakorda kirjete kogu, mille omadused (jaotused, korrelatsioonid) sarnanevad algsele andmestikule.

Nagu iga andmeanalüüs, algab ka süntees andmete ettevalmistamisega. Sünteesitud andmete kvaliteet sõltub sisendandmete kvaliteedist. Andmeteadlane puhastab andmed ning lingib andmestikud. Andmete linkimine pärast sünteesi ei ole ilma erilahendusteta võimalik.

Lähteandmeid kirjeldava masinõppemudeli loomiseks vaadeldakse, millisest jaotusest on erinevad tunnused ning millised tunnused on omavahel seotud ja kuidas. Oluline on jälgida, et masinõppemudel ei oleks ülesobitatud, sest muidu võib ta ”mäletada”isikustatud lähtenandmeid. Klassikaliste jaotuste puhul kasutatakse andmete genereerimiseks näiteks Monte Carlo meetodeid, mitteklassikaliste jaoks jõumeetodit, mille puhul luuakse juhuslikud punktid ning vaadatakse, kas need sobivad jaotusesse.

Pärast andmesünteesi on soovitav uurida loodud baasi kasulikkust ja privaatsust. Kasulikkus näitab, kui sarnane on genereeritud andmestik algsele andmestikule. Kõrge kasulikkus on hea, kui on vaja täpsust, näiteks, kui sünteetiliste andmeid on vaja kasutada masinõppes. Madal kasulikkus on vastuvõetav näiteks siis, kui on vaja infosüsteeme testida. Kasulikkust saab hinnata näiteks analüüsides mõlemaid andmebaase ning võrreldes tulemusi. Kui on juba ette teada, mis tüüpi analüüse hakatakse läbi viima, siis on võimalik andmestikke võrrelda just nende analüüside tulemuste põhjal. Üldiselt aga sünteesitakse andmeid, et andmestikul oleks n-ö sünteetiline kaksik, ning harva on ette teada, milliseid analüüse nende andmete peal täpselt tegema hakatakse (kui see oleks teada, siis võiks sünteesi asemel kohe analüüsi ära teha).

Teine võimalus on mõõta üldist kasulikkust, vaadeldes andmestike vahelist kaugust jaotuste, keskmiste ja standardhälvete abil. Selle meetodi keerukuseks on määrata ära, mis on piisav ja mis liiga suur erinevus andmestike vahel. Võimalik on ka kasutada subjektiivset hinnangut, kus eksperdid vaatavad peale andmestikule, kus on andmed nii algsest kui sünteesitud andmestikust, ning hindavad, millised on sünteesitud. Viimane meetod ei ole väga kestlik. Hetkel on kõige paremat kasutust leidnud kombineeritud meetod, kus kasutatakse kombinatsiooni analüüsitulemuste võrdlemisest ja andmestiku üldisest kasulikkusest.

Andmestiku privaatsuse mõõtmiseks on esmajoones võimalik võrrelda loodud kirjeid algsete kirjetega, et teha kindlaks, et sünteesitud andmed ei oleks samad, mis sisendandmed. Privaatsuse tagamiseks on oluline vältida masinõppemudeli ülesobitamist, sest vastasel juhul võib andmesüntees väljastada eriti võõrväärtustele liiga lähedasi väärtusi.

Sünteesitud andmeid kasutatakse et luua esialgseid masinõppemudeleid, kui ei ole võimalik saada juurdepääsu pärisandmetele[35]. Sünteesitud andmed kasutatakse ka tarkvara testimisel ja häkatonidel.

EU-SILC on üle-Euroopaline teadusprojekt vaesumisriski ja sotsiaalsete probleemide uurimiseks Euroopa Liidus. Projekti käigus sünteesiti pärisandmete põhjal mikroandmed [57].

[35] Simulacrum sisaldab Inglismaa riikliku vähiregistri andmete põhjal sünteesitud andmeid, mille põhjal saab kirjutada näidispäringuid, mis hiljem töötavad ka päris andmete peal. https://simulacrum.healthdatainsight.org.uk (viimati külastatud 27.02.2023).

**Turvagarantiid ja jääkriskid.** Sünteetilised andmed ei kuulu IKÜMi käsitlusalasse, kui kirjeid ei ole võimalik seostada algsete andmesubjektidega. Seega, kui masinõppemudel või neurovõrk on korralikult loodud ja genereeritud andmete kasulikkust ja privaatsust on kontrollitud, ei ole sünteesitud andmed pärisandmed. Seetõttu on neid võimalik kasutada infosüsteemide testimiseks ja andmeanalüüsi mõistlikkuse uurimiseks ilma pärisandmeid jagamata.

Oluline on seejuures tähele panna, et andmete sünteesile eelnev andmete puhastamine ja mudeli loomine on IKÜMi mõistes andmete töötlemine.

**Juhised rakendajale.** Andmete puhastamine ja süntees võib toimuda kõik ühes asutuses, aga on võimalik ka teenusena sisse osta nii osaliselt kui täielikult. Seega võib näiteks asutus andmed ise puhastada, aga sünteesi võib läbi viia keegi teine, aga asutus võib ka andmete puhastamise teenusena sisse osta. Nagu mainitud, on andmete sünteesile eelnev andmete puhastamine ja mudeli loomine IKÜMi mõistes andmete töötlemine, seega peab rakendajal olema õigus andmeid sellel eesmärgil töödelda.

**Õiguslikud aspektid.** Sünteetiliste andmete terminit seadusega defineeritud ei ole. Sünteetilisi andmeid kutsutakse vahel ka ”võltsandmeteks”(ingl ”fake data”) või kunstlikud andmed (ingl ärtificial data”). Sünteetilised andmed on fundamentaalsel tasandil algandmetest kunstlikult genereeritud andmed, mis säilitavad nimetatud algandmete statistilised omadused [58].

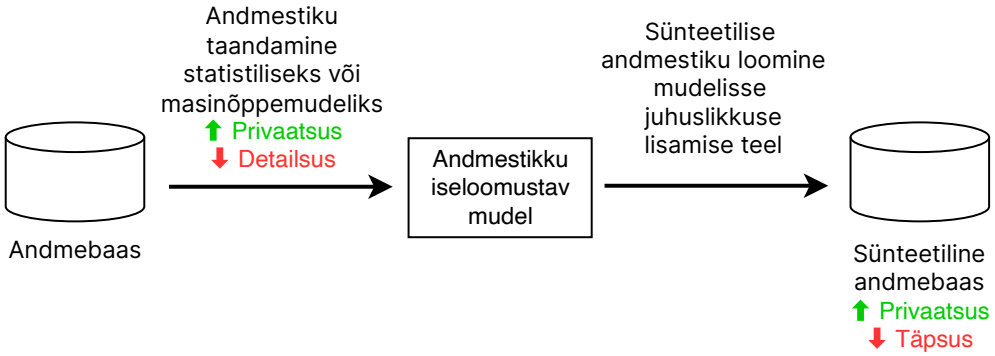
Sünteetilised andmed võivad põhineda nii isikuandmetel kui muudel andmetel, nt statistika. Kui sünteetilised andmed põhinevad isikuandmetel, siis tuleb selliste andmete töötlemisel järgida jurisdiktsioonis kehtivaid privaatsus- ja andmekaitse nõudeid. Kui isikuandmetel põhinevatel sünteetilistel andmetel ei ole võimalik ühtki sünteetilist andmepunkti algandmetele tagasi suunata, võivad olla tulemuseks sellised sünteetilised andmed, mis võivad jääda väljaspoole IKÜMi kohaldamisala. Vaatamata sellele tuleb andmekaitse nõudeid siiski järgida andmetöötluse faasides, mis tehakse enne sellise tulemuse saavutamist.

Sünteetiliste andmete puhul on välja kujunenud erinevad koolkonnad. Sünteetiliste andmete toetajad väidavad, et kui sünteetilised andmed on õigesti genereeritud, saavutavad need hästi seostamatuse eesmärki, st sünteetiliste kirjete seostamine inimesega on võimatu. Seetõttu käsitlevad mõned sünteetilisi andmeid anonüümsete andmetena. Sünteetiliste andmete vastased väidavad, et isegi siis, kui need on õigesti genereeritud, on üks-ühele seoseid endiselt võimalik luua, eriti kui sünteetiline andmekogum säilitab suure täpsusega algse andmekogumi omadused ja/või esinevad statistilised kõrvalekalded. Nendele eeldustele tuginedes peavad nad sünteetilisi andmeid tuvastatavaks teabeks [58].

Õiguslikust vaatenurgast võivad sünteetilised andmed pakkuda teatud juhtudel tõhusat kaitset isikuandmetele, mistõttu peetakse neid isikuandmete töötlemise paljulubavaks alternatiiviks. Peamine argument on see, et andmete sünteesi saaks kasutada tõhusa anonüümseks muutmise tehnikana andmetele juurdepääsuks, analüüsiks, jagamiseks, taaskasutamiseks ja avaldamiseks ilma isikuandmeid avaldamata. Andmete sünteesi peetakse sel määral vahendiks, mis järgib andmekaitsenõudeid ja stimuleerib samas tehnoloogilist innovatsiooni [58].

| **Sünteetiliste andmete genereerimine** | **ANDMED** |
|---|---|
| **Inglise keeles:** synthetic data generation | |
| **Lühidalt:** Sünteetilised andmed on juhuslikud andmed, mis näevad välja nagu pärisandmed. | **Arenduse keerukus:** keskmine |
| | **Ülalpidamise keerukus:** madal |
| | **Täpsus:** ebatäpne (andmed on statistiliselt sarnased, aga täpsus oleneb sünteesitud andmete kasulikkusest) |
| | **Privaatsusgarantii:** statistiline |
| | **Tehnoloogia küpsus:** keskmine |

**Ülevaatlik mudel:**




**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: Kui sünteesimiseks kasutatakse masinõppemudelit, peab rakendaja kindlaks tegema, et see mudel ei ole ülesobitatud.
2. Turvaeeldus: Rakendaja peab lisaks tulemuse kasulikkusele uurima ka tulemuse privaatsust.
3. Jääkrisk: mudelist genereeritakse juhuslikult päris inimene.

**Rakendusvõimalused:**

1. Andmete avaldamine testimiseks, õppe- ja teadustööks
2. Avaandmete avalikustamine ja teenused
3. Andmestike täiendamine, kui on puuduvaid väärtuseid või andmeid ei ole piisavalt

**Õiguspraktika:**

1. Kirjanduses on leitud, et sünteetilised andmed võiksid olla anonüümsed, kuid konkreetsed pretsedendid puuduvad.

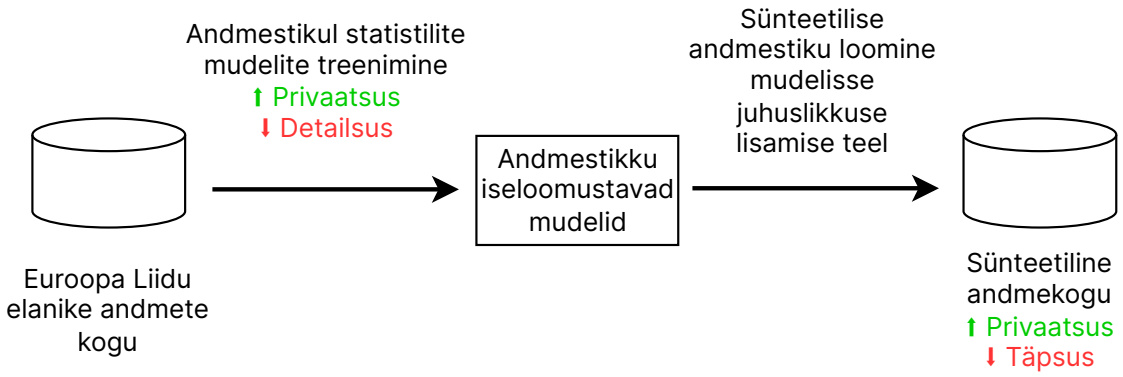
**Tuntumad rakendused:**

1. EU-SILC - Euroopa Liidu rahvastikuandmete sünteesimine
2. USA inimkaubandusevastane koostöö
3. Inimeste näopiltide süntees (This Person Does Not Exist teenus)

# EU-SILC – Euroopa Liidu rahvastikuandmete sünteesimine

| | |
|---|---|
| **Lühidalt:** EU-SILC projekt sünteesis sotsiaalprobleemide uurimiseks andmestiku Euroopa rahvastiku kohta. | **Teostamise aasta:** 2018 |
| | **Riik:** Euroopa Liit |
| | **Omanik:** Euroopa Liit |
| | **Teostaja:** Eurostat |
| | **Süsteemi küpsus:** püsiv juurutus |
| **Privaatsuskaitse tehnoloogiad:**<br>1. Sünteetiliste andmete genereerimine | **Sobivad kasutusjuhtumid:**<br>avaandmed |

**Ülevaatlik mudel:**




**Märkimisväärsed omadused:**

1. Sünteesiti Euroopa Liidu liikmesriikide andmeid kümne aasta ulatuses (2004-2013).
2. Koguti andmeid elanike ja leibkondade sissetulekute ja elatustaseme kohta Euroopa Liidu riikide elanikelt. Sünteesitud andmed genereeriti, et võimaldada soovijatel nende peal läbi viia statistilisi analüüse või teha esialgseid analüüse enne pärisandmetele ligipääsu küsimist,

### 4.2.8 Usaldatavad täitmiskeskkonnad

**Lihtsustatult.** Usaldatav täitmiskeskkond on arvuti, mis ei näe andmeid, mida ta töötleb ja ei saa seega neid lekitada. Usaldatava käivituskeskkonna korrektset tööd saab üle arvutivõrgu kontrollida.

**Ülevaade ja rakendamine.** Usaldatava täitmiskeskkonna tehnoloogia võimaldab arvutusseadme (olgu see telefon, tahvel või ka server) sees käivitada arvutusi ja andmetöötlust, mis on isoleeritud teistest sama seadme peal töötavatest arvutustest. Sisuliselt saab kaitsta arvuti sees töödeldavaid andmeid ka arvutit füüsiliselt kontrolliva isiku eest. Ehk siis - ehitatud on arvuti, mis ei saa lekitada enda töödeldavaid andmeid.

Sellisest arvutist on aga vähe kasu, kui sinna paneb kaitset nõudvaid andmeid juurde selle sama arvuti haldaja. Seega muutusid usaldatavad täitmiskeskkonnad privaatsuse jaoks huvitavaks siis, kui neile lisandus kaugatesteerimise võimalus. Atesteerimine lubab luua turvalise sidekanali mõne teise seadme ja isoleeritud käivitusskeskkonna vahel. See tähendab, et väline osapool saab eemalt ühendudes tõestuse, et ta suhtleb just usaldatava käivituskeskkonnaga ning saab seejärel sinna laadida töötlemiseks andmeid. Kui andmeid on vaja koguda rohkem või üle pikema perioodi, saab need salvestada käivituskeskkonnas hoitava võtmega ka salvestusseadmetele.

**Ajaloost** Usaldatavate täitmiskeskkondade (*Trusted Execution Environment*, TEE) tehnoloogia juured on võtmete ja saladuste kaitse tehnoloogiate juures. Trusted Computing Group eestvedamisel standarditi usaldatava platvormi moodulid (*Trusted Platform Module*, TPM), mida kasutatakse ennekõike krüptograafiliste võtmete riistvaraliseks kaitseks. Tänaseks on TPM tehnoloogia arvutites laialt levinud.

Aegamööda arendati tehnoloogiat edasi, et kaitsta saaks keerukamaid andmestruktuure ning ka nende üldotstarbelist töötlemist. Valminud tehnoloogiatest on teistest laialdasema levikuni jõudnud ARM-tüüpi arhitektuuriga protsessoritel saadaolev TrustZone tehnoloogia ning protsessoritootja Intel toodetavad Software Guard eXtensions (SGX) ja Trust Domain eXtensions (TDX) tehnoloogiad. Oma versioonid tehnoloogiast on olemas ka AMD, IBMi ja RISC-V tüüpi protsessorite jaoks.

**Turvagarantiid ja jääkriskid.** Usaldatavate käivituskeskkondade abil saab ehitada süsteeme, mis pakuvad riistvaralist-krüptograafilist konfidentsiaalsust töödeldavatele andmetele ning riistvaralist-krüptograafilist terviklust andmetele ja täidetava algoritmi koodile ning turvapoliitikatele. Kaugatesteerimise mehhanismi abil on võimalik garantiide kehtivust ka üle võrguühenduse kaugelt kontrollida.

Nende võrdlemisi tugevate turvagarantiide eelduseid on mitu:

1. rakenduse korrektne teostus (vt täpsemalt allpool),
2. protsessoritootja väljastatud uuenduste paigaldamine,
3. protsessoritootja (või vastava pilvandmetöötluse teenuse pakkuja) usaldusväärsus atesteerimisprotsessi ühe vahendajana ning
4. protsessoritootja mõningane usaldusväärsus tehnoloogia loojana.

Tavakasutajal on pea võimatu riistvaraliste turvalahenduste garantiisid kontrollida – tarkvara saab auditeerida, riistvara mitte nii väga. Õnneks. saab tavakasutaja loota selle peale, et infoturbekogukond tegeleb regulaarselt uute tehniliste turvatehnoloogiate uurimise ning haavatavuste avaldamisega.

Peamine teadaolev rünnete klass on nn kõrvalkanaliründed (*side channel attacks*). Selliste rünnete puhul püütakse mõõta arvutussüsteemi kõrvalkanaleid nagu näiteks tegevuste tööaega, energiakasutust, elektromagnetkiirgust, ning selle põhjal teha järeldusi arvutustes kasutatavate andmete kohta. Näiteks kui programmis tehakse andmepõhiseid otsuseid, on võimalik kõrvalkanaliründe abil teada saada, millist haru läbitakse, ning selle põhjal järeldada, mis oli mõne konfidentsiaalse andmeelemendi väärtus. Selliste rünnete vastu võitlemiseks on oluline usaldatavate käivituskeskkondade rakendusi hoolikalt arendada. Pikem analüüs Intel SGX turvamudeli ja kõrvalkanalirünnete kohta on vabalt saadaval [59].

Kõrvalkanalirünnete jääkriski kahandab hoolikas arendus, mis vähendab käivituskeskkonnas töötava koodi hulka. Tootjate usaldamise vajadust saab vähendada hoolikalt läbimõeldud võrguturbepoliitikaga ning kolmanda osapoole teenuste minimaalse rakendamisega.

**Juhised rakendajale.** Usaldavate käivituskeskkondade rakendamiseks on kolm peamist võimalust:

1. andmetöötlusrakenduse nullist teostamine tehnoloogia kasutust lihtsustavate arendusraamistike abil,
2. tehnoloogiat integreeritavate andmetöötlustoodete juurutamine ja
3. rakenduse käivitamine usaldatavat käivituskeskkonda kasutavas hüperviisoris.

Järgmises tabelis on selgitatud kolme lähenemise erinevusi.

| | | Automaatselt tagatud turvagarantiid | | | |
|---|---|---|---|---|---|
| Lahendus | Tooteid saadaval | Andmete privaatsus | Andmete terviklus | Rakenduse terviklus | Rollid ja õigused |
| Arendatud rakendus | Vähe | Jah | Jah | Jah | Jah |
| Analüütika-toode | Väga vähe | Jah | Jah | Piiratud | Ei |
| Virtualiseeri-mine | Vähe | Jah | Jah | Piiratud | Ei |

Rakenduste arendamise toetamiseks on mitmeid tehnoloogiaid (Google Asylo, Microsoft CCF, Sharemind HI), millest mõned toetavad rakenduste arendust üldiselt ja teised on mõeldud just privaatsust vajavate rakenduste loomiseks. Usaldatava käivituskeskkonna abil virtualiseerimist toetavad näiteks Anjuna ja Fortanixi tooted. Kuigi täna valmis analüütikatooteid ei ole (leidub arenduses prototüüpe), siis tehnoloogia on kiiresti arenemas ning on oodata, et lähiaastatel toodete valik kasvab kiiresti.

# Usaldatavad täitmiskeskkonnad

**ANDMED**

**Inglise keeles:** trusted execution environments

**Lühidalt:** Usaldatav täitmiskeskkond on arvuti, mis ei näe andmeid, mida ta töötleb ja ei saa seega neid lekitada. Usaldatava käivituskeskkonna korrektset tööd saab üle arvutivõrgu kontrollida.

**Arenduse keerukus:** keskmine

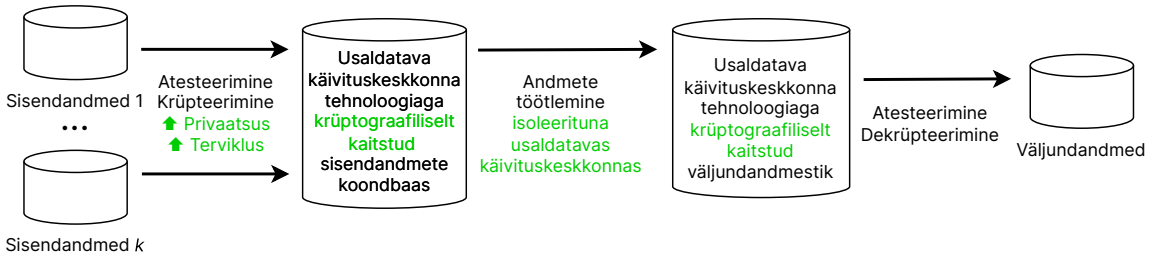
**Ülalpidamise keerukus:** madal

**Täpsus:** täpne

**Privaatsusgarantii:** riistvaraline

**Tehnoloogia küpsus:** keskmine

**Ülevaatlik mudel:**




**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: usaldatav täitmiskeskkond ja atesteerimine on korrektselt seadistatud.
2. Turvaeeldus: rakendus, mis keskkonnas töötab, on korrektselt teostatud ja kõrvalkanalirünnetele vastupidav.
3. Turvaeeldus: Tuleb usaldada täitmiskeskkonna tehnoloogiatootjat, et ta on teinud oma tööd korrektselt.
4. Jääkrisk: täitmiskeskkonna või rakenduse vea tõttu lekib konfidentsiaalseid andmeid.

**Rakendusvõimalused:**

1. Turvalised andmeruumid teenustele.
2. Andmete linkimise- ja analüüsiteenus.
3. Lisameetmena isikuandmete töötlemise kaitsel pilvandmetöötluses.
4. Tugevdava tehnoloogiana avaandmete ja päringusüsteemide teenustele.

**Õiguspraktika:**

1. Õiguslikke hinnanguid on vähe saadaval.

**Tuntumad rakendused:**

1. Kasutajate kontaktide ühisosa leidmine Signali vestlussüsteemis usaldatavate käivituskeskkondade abil.
2. Indoneesia Turisministeeriumi projekt sideandmete töötlemiseks.
3. Eurostati mobiilsusandmete longituuduuringu pilootprojekt.

### 4.2.9 Homomorfne krüptograafia

**Lühidalt.** Homomorfse krüptograafiaga kaitstud tundlikke andmeid saab töödelda nii, et neid selleks lahti ei krüpteerima ei pea.

**Ülevaade ja rakendamine.** Homomorfne krüptograafia on avaliku võtme krüptosüsteem, mis tähendab, et andmete krüpteerimiseks kasutatakse avalikku võtit ning dekrüpteerimiseks salajast võtit. Salajase võtme kaitsel on eriti suur tähtsus, sest igaüks, kellel on juurdepääs salajasele võtmele, saab lahti krüpteerida arvutuse vahe- või lõpptulemusi.

Homomorfsuse omadus tähendab, et sama võtmega krüpteerimisel saadud arvuliste väärtuste krüptotekste on võimalik omavahel kombineerida nii, et saadakse krüpteeritud arvude summa või korrutise krüptotekst. Selliseid toiminguid järjest kombineerides on võimalik teostada andmeid krüptograafiliselt kaitsvaid andmeanalüüsirakendusi.

Homomorfse krüptograafia juurutamise mudeli määrab paika avaliku ja salajase võtme kaitsevajadus. Avalikku võtit ei pea salajasena hoidma, seega andmeid saavad krüpteerida mitmed sisendit andvad osapooled. Küll aga peab igal tulemeid tarbival osapoolel olema koopia salajasest võtmest ning sellise võtme jagamine on võimalik turvarisk.

Eraldi kategooria moodustavad eriotstarbelised skeemid, näiteks määratud krüpteerimine (*deterministic encryption*), kus sama lähtetekst krüpteeritakse sama võtmega samaks krüptotekstiks ning järjestuslik krüpteerimine (*order preserving encryption*), kus krüpteeritud väärtused on omavahel samamoodi järjestatud nagu lähtetekstid. Neid on üritatud rakendada krüpteeritud andmebaasisüsteemide loomisel, kuid ei ole suudetud välistada lekkeid, mis tekivad kui analüüsida päringute mustreid ja andmete omavahelisi sõltuvusi koos [60].

**Ajaloost.** Tehnoloogia areng algas koos avaliku võtme krüptograafia arenguga ning üks tuntumatest homomorfsetest skeemid ongi RSA krüptosüsteem [61], mis lubab krüpteeritud väärtuseid omavahel korrutada. Rakendustes on levinud ka Paillier krüptosüsteem [62], mis lubab krüpteeritud väärtuseid omavahel liita. Vaid liitmisest või korrutamisest aga ei pane mõistlike andmetöötlusrakendusi kokku ning seega olid kirjeldatud skeemid kasulikud vaid eriotstarbeliste süsteemide (näiteks elektroonilised hääletussüsteemid) komponentidena.

Järgmine suurem läbimurre toimus 2009. aastal, kui leiutati esimene praktikas teostatav täishomomorfne (nii liitmist kui korrutamist toetav) krüptoskeemi [63]. Alguses olid sellised skeemid ääretult ebaefektiivsed, kuid aja jooksul on süsteemid efektiivsemaks muutunud ning jõutud on esmaste andmeanalüüsi ja isegi masinõpperakenduste katsetusteni.

**Turvagarantiid ja jääkriskid.** Homomorfsel krüptograafial põhinevate rakenduste peamised turvaeeldused on:

1. korrektne võtmehaldus arvutuse ajal ja järel,
2. alloleva krüpteerimisskeemi turvalisus ja
3. algoritmide kõrvalkanalikindlus.

Eelpool selgitasime, et homomorfse krüptograafiaga arvutatud tulemuste lahti krüpteerimiseks on igal vastaval osapoolel vaja salajast krüptograafilist võtit. Selle ühe võtme turvaline haldamine tähendab, et homomorfset krüptograafiat on keerukam efektiivselt juurutada üle mitme andmeid tarbiva organisatsiooni.

Uued täishomomorfsed krüpteerimisskeemid toetuvad keerukatele matemaatilistele ja keerukusteoreetilistele eeldustele, mida on uuritud, kuid mida veel lõpuni ei mõisteta ning seetõttu võib vähem uuritud skeemide kohta ilmuda ka uusi ründeid.

Nii nagu teiste turvalise arvutamise tehnoloogiatega, peab ka homomorfse krüptograafia teostusel jälgima, et privaatsed andmed ei lekiks tööaja kaudu.

Määratud krüpteerimise ja järjestusliku krüpteerimise rakendamisel tuleb arvestada võimalike ründeid paljude päringutega. Nende tehnoloogiate rakendamisele peab eelnema lahenduse põhjalik turvaanalüüs.

On väidetud, et olukorras, kus kasvõi ühel inimesel on juurdepääs andmete dekrüpteerimise võtmele,

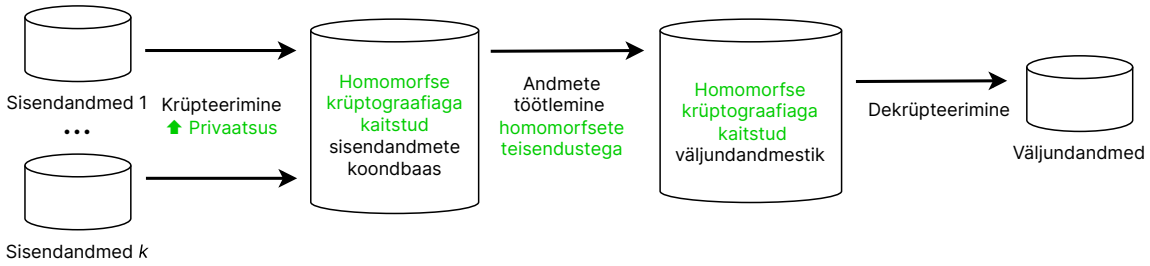
anonüümseid andmeid ei eksisteeri. Alternatiivne seisukoht on see, et kui vastutav andmetöötleja annab krüptitud andmed kolmandale osapoolele töötlemiseks ilma krüpteerimisvõtmeta, on sellel kolmandal isikul anonüümsed andmed. Näitena on toodud, et kui keegi edastab ajakohase ja piisava krüpteeringuga kaitstud andmed nii-öelda mustas kastis, siis võib eeldada, et mõistlikke vahendeid kasutades ei ole andmetöötlejal võimalik suletud mustas kastis olevaid andmeid töödelda [36].

**Juhised rakendajale.** Täna saab homomorfset krüptograafiat rakendada mitmete teekide abil. On ka iduettevõtteid, mis loovad tugilahendusi arendajatele. Täna vajab edukas homomorfse krüptograafia rakendamine siiski eraldi teadus-arendusprojekti.

[36] X-eHealth, D4.2.1 – Information paper on the current challenges in legal aspects of cross-border exchange of personal data. WP4 - Generic aspects of EEHRxF recommendation 27-05-2021. Version 0.7. Internetis kättesaadav: https://www.x-ehealth.eu/wp-content/uploads/2022/01/D4.2.1-–-Information-paper-on-the-current-challenges-in-legal-aspects-of-cross-border-exchange-of-person pdf (03.03.2023).

| **Homomorfne krüptograafia** | **ANDMED** |
|---|---|
| **Inglise keeles:** homomorphic encryption | |
| **Lühidalt:** Homomorfse krüptograafiaga kaitstud tundlikke andmeid saab töödelda nii, et neid selleks lahti ei krüpteerima ei pea. | **Arenduse keerukus:** kõrge<br>**Ülalpidamise keerukus:** kõrge<br>**Täpsus:** täpne<br>**Privaatsusgarantii:** tõestatav<br>**Tehnoloogia küpsus:** madal |

**Ülevaatlik mudel:**




**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: tagada korrektne võtmehaldus arvutuse ajal ja järel.
2. Turvaeeldus: veenduda alloleva krüpteerimisskeemi turvalises seadistuses.
3. Turvaeeldus: teostada algoritmid kõrvalkanalikindlalt.
4. Jääkrisk: teostuse vigade tõttu lekib konfidentsiaalseid andmeid.

**Rakendusvõimalused:**

1. Turvalised andmeruumid teenustele.
2. Lisameetmena isikuandmete töötlemise kaitsel pilvandmetöötluses.
3. Andmete linkimise teenus.

**Õiguspraktika:**

1. Õiguslikke hinnanguid on vähe saadaval.

**Tuntumad rakendused:**

1. Nõrkade paroolide tuvastamine Microsoft Edge veebilehitsejas homomorfse krüptograafia abil.
2. Šveitsi personaalmeditsiini võrk (koos teiste tehnoloogiatega).

### 4.2.10 Turvaline ühisarvutus

**Lühidalt.** Turvaline ühisarvutus aitab mitme osapoole saladustest arvutada uut teadmist ilma, et keegi teiste saladusi näeks.

**Ülevaade ja rakendamine.** Turvaline ühisarvutus on üldine tehnoloogia, millega saab ehitada mitmesuguseid rakendusi – seda kasutatakse nii andmete kui ka võtmete ja varade kaitseks. Andmete kaitsmise rakendustes on eesmärgiks kas isikute privaatsus või organisatsioonide saladuste kaitse. Võtmete ja varade kaitse puhul tehakse turvalise ühisarvutuse abil näiteks krüptograafilisi operatsioone või ka tehinguid digitaalsete varadega.

Andmetega töötavatest turvalise ühisarvutuse süsteemides eristatakse kolme tüüpi osapooli [64]. Sisendit andvad osapooled (*input party*) annavad salajasi andmeid ning soovivad neid salajasena hoida. Arvutavad osapooled (*computing party*) rakendavad turvalise ühisarvutuse protokolle, et andmed töödelda ilma neid nägemata. Tulemeid tarbivad osapooled (*result party*) näevad töötlemise väljundit, kuid ei midagi muud. Reaalsetes süsteemid võivad rollid kattuda – sisendit andvad osapooled võivad olla ka arvutavad ja väljundit tarbivat osapooled. Yao miljonäride probleemi lahendustes ongi kaks osapoolt, kes on kõigis kolmes rollis.

Privaatsuse tagamisel on turvaline ühisarvutus efektiivne ja sobilik olukordades, kus andmed tulevad mitmest allikast ning nende turvaline töötlemine ja kaitsmine on mitme osapoole huvides. Näiteid tehnoloogia erinevatest juurutusmudelistest saab lugeda UaESMC projekti tehnilisest aruandest [65]. Neist on ennast praktikas elujõulisemana näidanud kahe osapoolega arvutusmudelid inimeste ja asutuste vahel ning lihtsalt asutuste vahel. Need rakendused on modelleeritud klient-server mudelite järgi ning seega on sobinud hästi täna levinud juurutusmudelitesse.

Turvalise ühisarvutuse tugevused tulevad välja rohkemate osapooltega süsteemides, kuid täna veel puudub mitmel pool tehniline-organisatsiooniline küpsus, et seda täiel võimsusel rakendada. Seda põhjustab ennekõike nõue, et arvutavad osapooled peavad olema omavahel sõltumatud ning selleks vajalikku majutustaristut alles arendataks. Organisatoorsed ning tehnilised lahendused on küpsemas ning neid on juurutamas ja standardimas.

Sobivateks rakendusvaldkondades on koostööd ja isikustatud andmete nõudvad süsteemid. Tehnoloogia rakendusvõimaluste ning jõudluse kohta on avaldatud ka ülevaateartikkel [66].

Eestis on turvalist ühisarvutust rakendatud IKT riikliku programmi pilootprojektis PRIST (Privacy-preserving statistical studies on linked databases). Projekti raames tehti turvalise ühisarvutusega ulatuslik privaatsust säilitav isikuandmete linkimine ja statistiline uuring Maksuameti ja Haridus- ja Teadusministeeriumi andmetel. Uurimuse eesmärk oli mõista seoseid ülikooli ajal töötamise ja õigeaegselt lõpetamise vahel [67, 68]

**Tehnoloogia saamisloost.** Turvalise ühisarvutuse (*secure multi-party computation*) tehnoloogia aluseks peetakse Andrew Chi-Chih Yao 1986. aasta artiklit [69], milles ta sõnastab enda järgi nimetatud Yao miljonäride probleemi - kuidas saavad kaks miljonäri teada, kumb on rikkam ilma enda enda varanduse mahtu avaldamata? Turvaline ühisarvutus lahendabki kahe ja rohkema osapoole jaoks ülesandeid, kus osapoolte sisend jääb salajaseks kõigi teiste eest. Tehnoloogia oli teoreetiline kuni 2004. aastani, mil lisraeli teadlased ehitasid Fairplay prototüübi ja näitasid esimest korda turvalist ühisarvutust praktikas [70]. Pärast seda hakkas tehnoloogia kiiresti arenema ning esimesed reaalseid andmeid kasutanud rakendused ehitati Taanis [71] ja Eestis [72].

**Turvagarantiid ja jääkriskid.** Turvalise ühisarvutuse definitsiooni järgi ei tohi keegi peale sisendit andnud osapoole näha selles sisendi väärtuseid (kui see pole mingil põhjusel kokku lepitud kui teatud tingimustel lubatud väljund). Praktikas tähendab see, et korrektselt ehitatud turvalise ühisarvutuse süsteemis saavutatakse otsast otsani krüpteerimine andmeanalüüsile. Andmete omanik rakendab oma andmetel krüpteerimist (või midagi analoogset, nt ühissalastust) ning edastab andmed töötlemiseks. Töötlemine toimub krüpteeritud andmetel ilma neid lahti krüpteerimata ning väljund on samuti krüpteeritud kujul. Seega on turvalisus võrreldav arvutivõrkude turvakanalitega, kus kliendi ja serveri vahel ei ole keegi võimeline edastatud andmeid lugema.

Turvaline ühisarvutus üksi annab privaatsust otsivale andmeanalüüsisüsteemile matemaatiliselt tõesta-

tava krüptograafilise turvalisuse. Päringute piirangute teostamiseks on turvaline ühisarvutus samuti sobiv – privaatsuspoliitika ning päringute piirangute kehtestamise garantii on turvalise ühisarvutuse puhul väga efektiivne tänu hajusale ja konsensulikule kontrollile. Allikaprivaatsus ja väljundiprivaatsus on teostatavad rakenduse tasemel.

Turvalise ühisarvutuse rakenduste turvaeeldused on:

1. krüptograafilise protokollistiku turvaeelduste täitmine (sh osapoolte sõltumatus) ja
2. käivitatavate algoritmide kindlus kõrvalkanalirünnete vastu.

Turvalise ühisarvutuse protokollide tavapärane turvaeeldus on, et kas enamus või vähemalt üks arvutavatest osapooltest käitub ausalt ning ei ürita protokolli rikkuda. Vastasel juhul suudaks mingi alamosa arvutavatest osapooltest enda käes olevaid krüpteeritud materjale kokku pannes mõned saladused avalikuks teha. Praktikas saavutatakse seda nii tehniliste kui organisatoorsete meetmetega. Kõige tugevamaks riski kahandamise meetmeks on arvutavate osapoolte majutamine sõltumatute organisatsioonide või inimeste poolt. See tähendab, et turvalise ühisarvutuse süsteemi ei tohiks juurutada samas pilvarvutussüsteemis või andmekeskuses. Ühe organisatsiooni piires juurutamine ei ole välistatud, kuid siis peab tagama organisatsiooni sees selle, et serverite administraatorid koostööd ei tee.

Turvalise ühisarvutuse puhul tuleb samuti jälgida, et andmeanalüüsi algoritmi tööaeg ei lekiks privaatsete sisendandmete väärtuseid. Vastavat turvamudelit ning turvalise ühisarvutuse algoritmide koostamise meetmeid on kirjeldanud näiteks [73].

**Juhised rakendajale.** Turvalise ühisarvutuse tehnoloogiate pakkujaid on maailmas mitmeid ning neid ühendab ka tööstusliit MPC Alliance[37] Tehnoloogia on täna saadaval ennekõike valdkonnaspetsiifiliste valmisrakenduste või üldiste programmeeritavate raamistike kujul. Viimaste puhul on saadaval nii tööstuslikul tasemel teostuseid kui ka avatud lähtekoodiga akadeemilisi prototüüpe. Üldotstarbeliste turvalise ühisarvutuse raamistike küpsuse kohta on avaldatud ka ülevaateartikkel [74].

Täna tähendab turvalise ühisarvutuse rakendamine süsteemis arendusprojekti, mille käigus saab ära kasutada olemasolevaid tehnoloogiakomponente - st krüptograafiat uuesti leiutada või teostada ei ole vaja. Suuresti on tegemist integratsiooniprojektiga, kus oluline roll on äriloogika ja juurutusmudeli kavandamisel ja teostamisel vastavate arendusvahenditega.

**Õiguslikud aspektid.**

Õiguslikud arutelud selle üle, kas turvaline ühisarvutus tagab anonüümsust Isikuandmete kaitse üldmääruse mõistes, ei ole lõplikku tulemust andnud. Teaduskirjanduses on seda väidetud [75], kuid hilisem IKÜM praktika on olnud ebaselge. Turvalise ühisarvutuse protokolli osapooled on võimalik pidada kaasvastutavateks või alamtöötlejateks (sõltuvalt juurutusest), kes küll andmeid reaalselt isikustada ei suuda, kuid suunavad nende täitmist.

Siiski peetakse turvalist ühisarvutust tugevaks täiendavaks seostamatuse ja turvameetmeks. Euroopa Andmekaitsenõukogu oma 2020. aasta novembri arvamuses võtnud seisukoha, et turvaline ühisarvutus võib olla sobiv meede isikuandmete kaitseks nende transpordil väljaspoole Euroopa Liitu (vt [76], Lisa 2, kasutusjuhtum 5).

[37] MPC Alliance. https://www.mpcalliance.org (viimati külastatud 11.01.2023).

# Turvaline ühisarvutus

**ANDMED**

**Inglise keeles:** secure multi-party computation

**Lühidalt:** Turvaline ühisarvutus aitab mitme osapoole saladustest arvutada uut teadmist ilma, et keegi teiste saladusi näeks.

**Arenduse keerukus:** kõrge

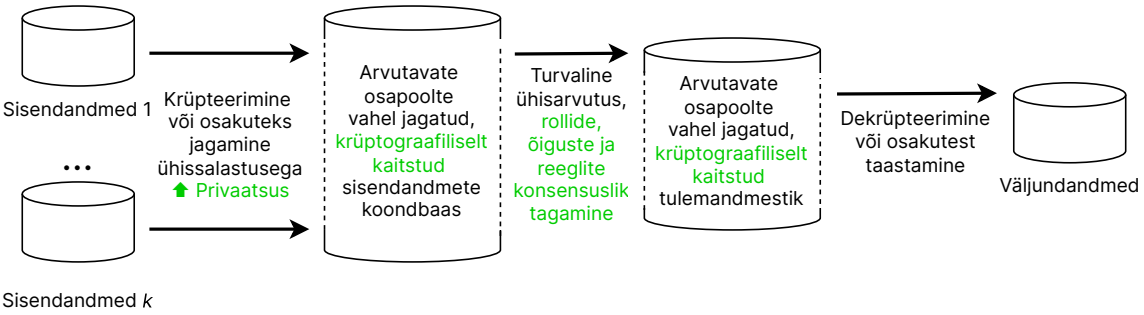
**Ülalpidamise keerukus:** kõrge

**Täpsus:** täpne

**Privaatsusgarantii:** matemaatiliselt tõestatav

**Tehnoloogia küpsus:** keskmine

**Ülevaatlik mudel:**




**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: tagada korrektne võtmehaldus.
2. Turvaeeldus: teostada algoritmid kõrvalkanalikindlalt.
3. Turvaeeldus: taga arvutavate osapoolte omavahelise sõltumatuse nõue, vajadusel lepingute toega.
4. Jääkrisk: teostuse vigade tõttu lekib konfidentsiaalseid andmeid.

**Rakendusvõimalused:**

1. Turvalised andmeruumid teenustele.
2. Lisameetmena isikuandmete töötlemise kaitsel pilvandmetöötluses.
3. Andmete linkimise- ja analüüsiteenus.
4. Tugitehnoloogiana avaandmete ja andmebaaside avaldamise teenustele.

**Õiguspraktika:**

1. Euroopa Andmekaitsenõukogu (EDPB) on lugenud tehnoloogia sobivaks lisameetmeks isikuandmete edastamisel väljaspoole Euroopa Liitu.

**Tuntumad rakendused:**

1. Bostoni linna palgalõhe uuring turvalise ühisarvutuse abil.
2. Eesti IKT erialade tudengite andmete ühendamine ja õpikäitumise uuring turvalise ühisarvutusega.
3. Suurbritannia valitsuse piloot toetuspettuste uurimiseks turvalise ühisarvutusega.

# Eesti IKT erialade tudengite õpikäitumise ja töötamise uuring

**Lühidalt:** Uuringu käigus ühendati turvalise ühisarvutuse abil isikukoodide järgi IKT tudengite töötulumaksu- ja hariduse andmed ning uuriti, kuidas on töötamine ülikooliõpingute ajal seotud nominaalajas lõpetamisega.

**Teostamise aasta:** 2013-2015

**Riik:** Eesti

**Omanik:** Turvalist arvutuskeskkonda haldasid Riigi Infosüsteemi Amet (RIA), Rahandusministeeriumi Infotehnoloogiakeskus (RMIT) ja Cybernetica.

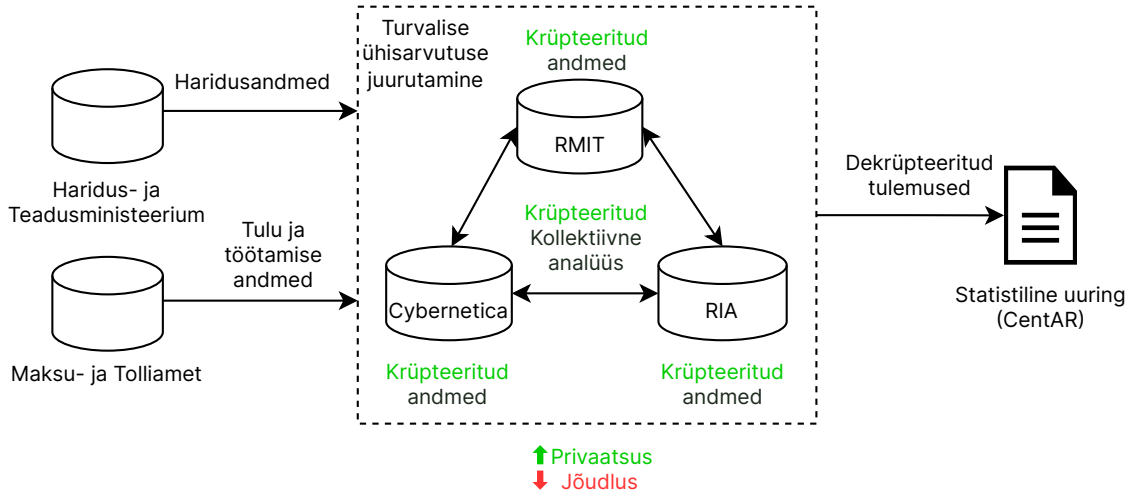
**Teostaja:** Statistilist analüüsi tegi Eesti Rakendusuuringute Keskus (CentAR)

**Süsteemi küpsus:** pilootprojekt

**Privaatsuskaitse tehnoloogiad:**
1. turvaline ühisarvutus

**Sobivad kasutusjuhtumid:**
turvaline linkimine ja analüütika

**Ülevaatlik mudel:**




**Märkimisväärsed omadused:**

1. Andmete töötlemine ja analüüs toimus Cybernetica poolt arendatud turvalise ühisarvutuse süsteemi Sharemind MPC abil. See kindlustas, et privaatsed andmed olid krüpteeritud kogu protsessi käigus ning avaldati ainult analüüside tulemused.
2. Kokku analüüsiti enam kui 10 miljonit kirjet Maksu- ja Tolliametilt ja enam kui 600 000 kirjet Haridus- ja Teadusministeeriumilt. Tegemist on ühe suurima krüpteeritud andmetega tehtud statistilise analüüsiga.

## 4.3 Identiteedi ja tõendustehingute kaitse

### 4.3.1 Pimesignatuurid

**Lühidalt.** Pimesignatuurid lubavad küsida serverilt signatuure andmetele, mida server näha ei saa. Kaitstakse digitaalselt allkirjastatava sõnumi privaatsust.

**Ülevaade ja rakendamine.** Pimesignatuur, õigemini selle väljastamine on kahe osapoole vaheline protokoll, mille käigus esimene osapool (klient) saab teiselt osapoolelt (serverilt) allkirja sõnumile, mille klient valinud on, aga signeerimise käigus ei saa server teada, millele ta allkirja andis. Signatuuri on võimeline verifitseerima suvaline kolmas osapool, kes teab serveri avalikku võtit. Tüüpiline on nõuda, et signeerimisprotokoll oleks üheraundiline – klient pimendab oma sõnumi, saadab pimendatud sõnumi serverile, server arvutab oma privaatvõtme abil pimendatud sõnumile signatuuri ja saadab selle kliendile tagasi, klient eemaldab signatuurilt pimenduse. Sõltuvalt konkreetsest krüptograafilisest konstruktsioonist võib olla võimalik täpsem kontroll selle üle, millised osad signeeritavast sõnumist serveri eest varjatuks jäävad.

Rõhutame, et kuigi server ei saa teada, millise sõnumi ta signeeris, kontrollib ta siiski allkirjastamise protsessi. Ilma serveri allkirjastamissoovita ei ole võimalik sõnumeid signeerida. Võimatus tekitada rohkem signatuure kui serveri signeerimissessioonide arv on pimesignatuuride peamine terviklusomadus.

**Ülevaade ja rakendamine.** Pimesignatuurid on ilmselt üks vanimaid krüptograafilisi privaatsustehnoloogiaid. Tehnoloogiat kirjeldas 1983. aastal David Chaum [77], olles aasta varem välja pakkunud paar võimalikku rakendust sellele tehnoloogiale [78]. Üks väljapakutud rakendus oli elektrooniline hääletamine, kasutades protokollistikku, mille üheks osaks on kõigi esitatud häälte avalikustamine, peale mida võib ükskõik kes nad kokku lugeda. Hääled ei sisalda informatsiooni neid esitanud isikute kohta, vastupidine oleks ränk privaatsusriive. Otsustamaks, milline hääl on kehtiv, nõuame aga, et valimisi läbiviiv asutus või mõni tema poolt volitatud isik peab olema hääled allkirjastanud. Hääle esitamise protokoll sisaldab seega hääletaja päringut valimisi läbiviivale asutusele, sooviga saada allkiri oma häälele. Asutus kontrollib hääletaja õigust hääletamisel osaleda ning signeerib hääle. Selleks, et asutus ei saaks teada, milline on allkirjastatav hääl, kasutame pimesignatuure.

Ilmselt tuntuim pimesignatuuride rakendus, mille David Chaum välja pakkus, on *digitaalne sularaha*. Seda mõistet on viimase neljakümne aasta jooksul ilmselt mitmeti sisustatud, kuid toona tähendas see elektroonilist maksesüsteemi, mille anonüümsusomadused on sarnased sularahaga. Selles süsteemis liigub teatud nominaalväärtusega rahatäht kõigepealt pangast välja, ostja kätte, seejuures debiteerib pank ostja kontot rahatähe väärtuse võrra. Ostja annab rahatähe üle müüjale, eeldatavasti saades vastu rahatähe nominaalväärtusega võrreldavas koguses kaupa või teenust. Müüja saadab rahatähe panka ning pank krediteerib müüja kontot. Pank ei tohi suuta välja antud ja tagasi tulnud rahatähti omavahel seostada. Süsteemi realisatsioonis on pangal rahatähtede iga võimaliku nominaalväärtuse jaoks avalik võti, millele vastava privaatvõtmega ta rahatähti allkirjastab. Rahatähe pangast saamiseks genereerib ostja juhusliku sisuga sõnumi ja laseb pangal selle signeerida, kasutades pimesigneerimist. Kui müüja saab ostjalt või pank müüjalt rahatähe, siis ta kontrollib, et pank on selle õige võtmega allkirjastanud. Hilisematel digitaalse sularaha süsteemidel võib olla mitmeid täiendavaid omadusi, nagu topeltkulutamise vältimine [79, 80, 81] (oluline, kui ostja ja müüja vaheline ning müüja ja panga vaheline protokoll toimuvad eri aegadel, mis võis olla tingitud sellest, et ostu-müügi hetkel töötasid ostja ja müüja vallasrežiimis), rahatähe osaline kulutamine [82], võimalus anda rahatäht üle teisele ostjale [83, 84] jne.

Digitaalse sularaha tehnoloogiate, sh. pimesignatuuride kommertsialiseerimiseks rajas David Chaum 1989. aastal ettevõtte DigiCash[38]. Ettevõttel ei õnnestunud oma tehnoloogiate kasutajaskonda kasvatada ning 1998. aastal läks see pankrotti[39].

Tänapäeval on pimesignatuuride kaks põhilist kasutamisviisi põhimõtteliselt needsamad kaks, mida kirjeldas David Chaum oma esimeses artiklis [78]. Pimesignatuure võib kasutada loatähtede loomiseks, kus loatähe identiteet (seerianumber) on mingi suvaline sõna, kuid oluline on, et õige autoriteet on tema all-

[38] https://en.wikipedia.org/wiki/DigiCash (viimati külastatud 01.03.2023).

[39] Digicash files Chapter 11 https://www.cnet.com/tech/tech-industry/digicash-files-chapter-11 (viimati külastatud 01.03.2023).

kirjastatunud. Pimesi loodavaid loatähti kasutavad nii Google One VPN[40] kui ka Apple'i iCloud Private Relay[41], kus kasutusõigust kontrolliva serveri allkirjastatud loatäht on vajalik teenusele ligipääsuks. Schmitt ja Raghavan [85] kirjeldavad mobiilivõrgu arhitektuuri, kus telefoni loatäht teenusega ühendumiseks on pimesigneeritud, varjates signeerija eest kliendi ja telefoni unikaalseid identifikaatoreid. Märgime, et VPN-teenuste puhul on võimalik vältida kasutaja jälitamist kolmandate osapoolte poolt, aga VPN teenuse pakkujat ennast tuleb endiselt usaldada [86].

Pimesignatuurid on kasutatavad ka anonüümsust mittepakkuvate plokiahelapõhiste krüptorahaplatvormide privaatsusomaduste parandamiseks [87].

Pimesignatuure võib kasutada ka tõendamaks, et mingid andmed on liikunud läbi mingi kindla osapoole, seejuures varjates neid selle osapoole eest. Meta Inc. soovib pettustega võitlemiseks koguda kliendiseadmetest teatud andmeid, et neid siis üheskoos analüüsida. Nad soovivad, et andmekogusse jõuaksid ainult legitiimsetelt klientidelt pärit andmed, vältimaks otsuste tegemist halbade andmete alusel. Nad on välja pakkunud arhitektuuri, kus Meta taristu komponendid signeerivad kliendi logikirjeid, kui kliendiseade nende komponentidega ühendub. Signeerimine toimub pimesi[42].

Pimesignatuuride standardiseerimine leiab hetkel aset läbi IETF-i[43]. Olemasolevale standardimustandile on olemas krüptoanalüüs, mis ta turvaliseks tunnistab [88]. Standardimustand pakub välja sellised signeerimisprotokollid ja andmeformaadid, et tulemusena tekib standardne RSA-PSS signatuur[89], seega võiks mingite serverite väljastatavate pimesignatuuride lisamine infosüsteemidesse, kus kasutajate privaatsust neist võidab, üsna lihtne olla. Pimesignatuuride juurutamist hõlbustab ka see, et nende konstruktsioonid ei vaja „eksootiliste" algebraliste struktuuride ja operatsioonide, nagu näiteks bilineaarsed paarimised, kasutamist.

**Turvagarantiid ja jääkriskid.** Peamine pimesignatuuride privaatsusgarantii on, et allkirjastav teenus ei näe allkirjastatavat sõnumit. Peamine terviklusgarantii on, et ühe teenuse poolt ei saa olla olemas rohkem allkirjastatud dokumente kui on selle teenuse kasutuskordade arv.

Teatud rakendustes on vaja täiendavaid võimalusi. Näiteks digitaalse raha skeemides topeltkulutamise vältimine toetub võimalusele mingil viisil ikkagi teada saada, millal sõnum signeeriti. Anonüümsuse kõrvaldamine on täiendav operatsioon pimesignatuuridel, mida mõni konstruktsioon toetab ja mõni mitte. Osades konstruktsioonides nõuab anonüümsuse kõrvaldamine usaldatud kolmandat osapoolt, mõnes teises suudab pank sõnumi signeerimishetke sõnumiga ise kokku viia, kui ta näeb tagasi tulnud rahatähte mitu korda.

**Juhised rakendajale.** Pimesignatuurid on komponent mingites suuremates infosüsteemides. Infosüsteemi arhitekt peab aru saama, kas pimesignatuurid sobituvad selle süsteemi nõuetega. Pimesignatuuride krüptograafia on põhjalikult uuritud ning standardimisel. Teostamist lihtsustavad avatud lähtekoodiga koodinäited ja teegid.

[40] VPN by Google One, explained https://one.google.com/about/vpn/howitworks (viimati külastatud 01.03.2023).

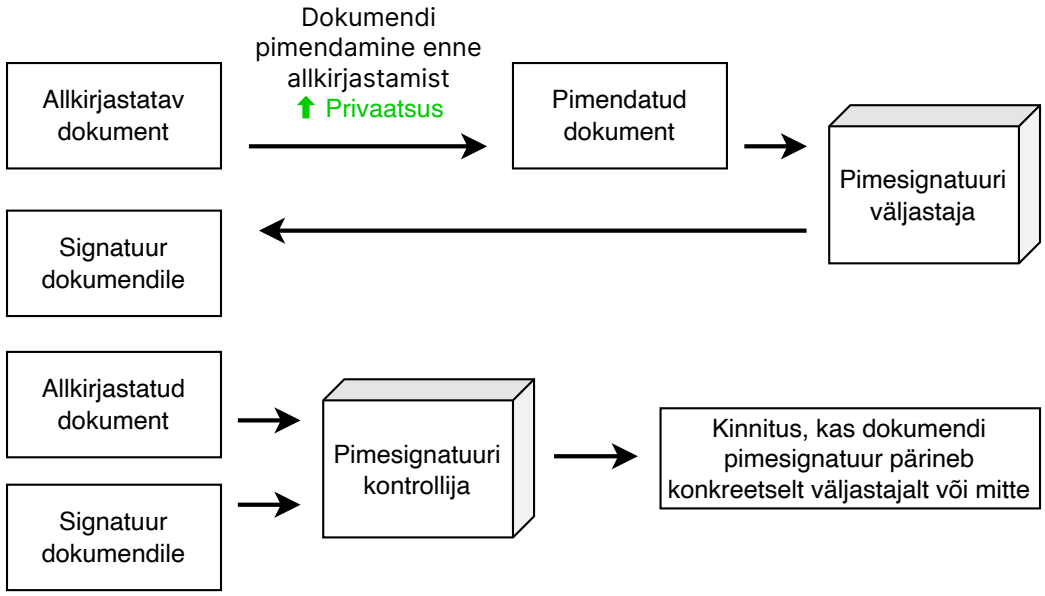
[41] iCloud Private Relay Overview https://www.apple.com/icloud/docs/iCloud_Private_Relay_Overview_Dec2021.pdf (viimati külastatud 01.03.2023).

[42] Fighting fraud using partially blind signatures https://engineering.fb.com/2019/10/16/security/partially-blind-signatures (viimati külastatud 01.03.2023).

[43] RSA Blind Signatures https://datatracker.ietf.org/doc/draft-irtf-cfrg-rsa-blind-signatures (viimati külastatud 01.03.2023).

| **Pimesignatuurid** | **IDENTITEET** |
|---|---|
| **Inglise keeles:** Blind signatures | |
| **Lühidalt:** Pimesignatuurid lubavad küsida serverilt signatuure andmetele, mida server näha ei saa. Kaitstakse digitaalselt allkirjastatava sõnumi privaatsust. | **Arenduse keerukus:** keskmine<br>**Ülalpidamise keerukus:** madal<br>**Täpsus:** —<br>**Privaatsusgarantii:** krüptograafiliselt tõestatav<br>**Tehnoloogia küpsus:** keskmine |

**Ülevaatlik mudel:**




| **Turvaeeldused ja jääkriskid:** | **Rakendusvõimalused:** |
|---|---|
| 1. Turvaeeldus: Serveri ja kliendi funktsionaalsus tuleb teostada korrektselt.<br>2. Jääkrisk: teostuse vea tõttu lekib konfidentsiaalseid andmeid. | 1. Eriotstarbelised süsteemid, mis vajavad oma töö käigus selliste andmete digitaalset allkirjastamist, mida allkirjastaja näha ei tohiks. |
| **Õiguspraktika:**<br>1. Õiguslikke hinnanguid on vähe saadaval. | **Tuntumad rakendused:**<br>1. Google One VPN<br>2. Apple iCloud Private Relay |

# Pimesignatuurid Google One virtuaalses privaatvõrgus

**Lühidalt:** Google One virtuaalne privaatvõrk (VPN) on teenus, mida Google pakub osana oma tellimuspõhisest teenusest Google One. Teenust eristab traditsioonilistest VPN teenustest, mis võivad kasutajate andmeid lekitada nende identiteedi ja võrguliikluse linkimise tõttu, pimesignatuuride kasutamine autentimise ja virtuaalse privaatvõrguga ühendamise juures.

**Teostamise aasta:** 2020

**Riik:** Ameerika Ühendriigid

**Omanik:** Google

**Teostaja:** Google

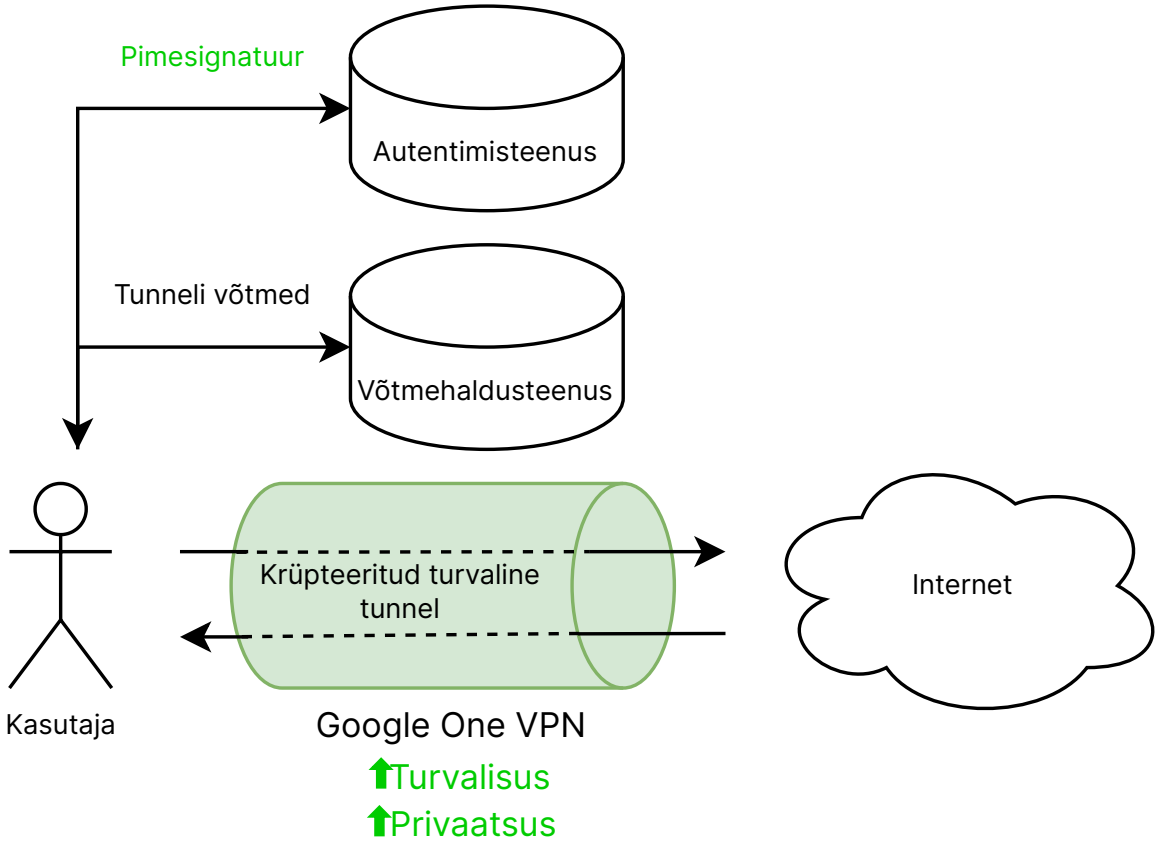
**Süsteemi küpsus:** püsiv juurutus

**Privaatsuskaitse tehnoloogiad:**

1. pimesignatuurid
2. otspunktkrüpteerimine

**Sobivad kasutusjuhtumid:**
anonüümne side privaatvõrgus

**Ülevaatlik mudel:**




**Märkimisväärsed omadused:**

1. Pimesignatuuride kasutamine pakub tugevamaid privaatsusgarantiisid ning raskendab privaatvõrgu serveritel kasutaja identiteedi seostamist kasutaja võrguliiklusega.
2. Google on teenuse jaoks arendanud vabavaralisi kliendirakendusi, mis on saadaval mitmel platformil, kuid ainult teatud riikides ja tellimuspõhiselt.
3. Meedias on mainitud, et kuigi tehnilise lahenduse poolest lubab Google One VPN paremat privaatsusgarantiid kui teised konkurendid, peab kasutaja usaldama, et tehnoloogiafirma ise andmeid ei kasutaks või kolmandatele osapooltele ei müüks.

### 4.3.2 Rühma- ja ringisignatuurid

**Lühidalt.** Rühmasignatuurid lubavad anda digitaalset allkirja grupi nimel. Rühmasignatuur varjab konkreetse allkirjastaja isikut.

**Ülevaade ja rakendamine.** Rühmasignatuurid lubavad mitmel isikul / autoriteedil moodustada *rühma* ja pärastpoole anda allkirju *rühma nimel*. See tähendab, et allkirja kontrollija veendub, et see on signeeritud selle rühma mõne liikme poolt, aga ta ei saa teada, millise liikme poolt. Võrreldes tavalise signeerimise ja signeerimisvõtme kõigile rühmaliikmetele andmisega, võimaldavad rühmasignatuurid allkirja anonüümsuse eemaldada, kui mingi hulk osapooli nii otsustab. Tüüpilised rühmasignatuuriskeemid on loodud nii, et selleks hulgaks osapoolteks on üksainus usaldatud osapool *rühmahaldur*, kes ise rühma liige ei ole.

Rühmasignatuuri primitiivi pakkusid välja Chaum ja van Heyst [90] 1991. aastal. Kasutusjuht, millega nad seda primitiivi motiveerisid, oli *anonüümne autentimine* – ühenduse loomisel saab server teada, et temaga on ühendust võtnud rühma liige, aga ei saa teada, milline. Hilisemad rakendused on olnud sarnased – rühmasignatuurid lubavad isikul rühma sees anonüümseks jäädes ennast autentida või dokumente allkirjastada. Osade kasutusjuhtude juures on oluline, et mingitel tingimustel on võimalik isik siiski tuvastada.

Rühmasignatuuridega sarnaste primitiivide kirjeldamiseks ja omavahel võrdlemiseks toome sisse mõned tähistused. Enne osapooltest $P_1, \ldots, P_n$ koosneva rühma $\mathcal{G}$ nimel signeerimisi peab olema toimunud *rühma loomise protokoll*, mille käigus luuakse rühma avalik võti $pk^{\mathcal{G}}$, iga liige $P_i$ saab oma signeerimisvõtme $sk_i^{\mathcal{G}}$ ning rühmahaldur saab vajalikud võtmed, millega on võimalik signeerija anonüümsus tühistada. Kui osapool $P_i$ soovib luua signatuuri $\sigma$ sõnumile $m$, siis kasutab ta oma võtit $sk_i^{\mathcal{G}}$. Signatuuri $\sigma$ verifitseerimiseks kasutatakse võtit $pk^{\mathcal{G}}$; signatuurist ei ole võimalik ilma rühmahalduri abita leida, millisega võtmetest $sk_1, \ldots, sk_n$ see leiti.

Kui rühmasigneerimisel on vajalik kõigepealt moodustada rühm, siis *ringisignatuurides* võib allkirjastaja igal signeerimisel uuesti valida, kes seekord kuuluvad rühma, mille sees ta anonüümseks jääb. Ringisigneerimisel on igal isikul oma avalik võti $pk$ and signeerimisvõti $sk$ nagu „tavaliselgi“ signeerimisel. Kui aga isik $P$ soovib signeerida sõnumit $m$, siis ta otsustab, milliste teiste isikutega $P_1, \ldots, P_k$ ta soovib rühma kuuluda. Signeerimine vajab lisaks sõnumile $m$ ja isiku $P$ signeerimisvõtmele $sk$ veel isikute $P_1, \ldots, P_k$ avalikke võtmeid $pk_1, \ldots, pk_k$. Loodud signatuuri $\sigma$ verifitseeritakse avalike võtmete hulga suhtes, millesse kuuluvad $pk_1, \ldots, pk_k$ ja $pk$ (mis vastab signeerimisvõtmele $sk$); verifitseerimisel ei selgu, millisele neist $(k+1)$-st avalikust võtmest vastava signeerimisvõtmega on $\sigma$ loodud. Leidub ringisignatuuride konstruktsioone, kus selline deanonüümimine on igal juhul võimatu. Samuti leidub konstruktsioone, kus mingi usaldatud kolmas osapool on suuteline signatuurist $\sigma$ ja verifitseerimisel kasutatavatest avalikest võtmetest $pk_1, \ldots, pk_k$ leidma, milline isik selle signatuuri lõi.

Esimese näite ringisignatuurist tõid Chaum ja van Heyst sellessamas artiklis, kus nad pakkusid välja rühmasignatuurid [90], kasutamata seejuures küll seda nime. Ringisignatuuri spetsifitseerisid Rivest, Shamir ja Tauman [91] 2001. aastal.

Ringisignatuuride signeerimisalgoritmides on võimalik eristada kahte sammu:

- „Esialgne“ signeerimine, mille käigus kombineeritakse sõnum $m$ ja signeerija signeerimisvõti $sk$. Samuti luuakse mingid väärtused, mida kasutatakse teisel sammul.
- Loodud signatuuri anonüümimine, mille käigus kombineeritakse esimese sammu tulemus avalike võtmetega $pk_1, \ldots, pk_k$ ning saadakse ringisignatuur $\sigma$.

Eri konstruktsioonides võivad need sammud rohkem või vähem kombineeritud olla, aga kontseptuaalselt on võimalik need kaks sammu teineteisest eraldada. Enamgi veel: pole põhjust nõuda, et anonüümimiseks oleks tarvis signeerimisvõtit $sk$. Nii jõuame *anonümiseeritavate signatuurideni*. Neil signeerimisskeemidel moodustatakse signeerimisoperatsiooni ajal tavaliste omadustega signatuur $\sigma^\flat$, mida on põhimõtteliselt võimalik signeerija avaliku võtme $pk$ suhtes verifitseerida. Hiljem on võimalik välja valida rühm isikuid $P_1, \ldots, P_k$, kelle seas me signeerija anonüümsust soovime tagada, kombineerida $\sigma^\flat$ vastavate avalike võtmetega $pk_1, \ldots, pk_k$ ja saada tulemuseks ringisignatuur $\sigma$, mille verifitseerimise käib taas avalike võtmete hulga $pk_1, \ldots, pk_k, pk$ suhtes. Rühma välja valija ja ringisignatuuri arvutaja ei pea olema sama isik kui see, kes signatuuri lõi. Anonümiseeritavate signatuuride primitiivi ja esimese konstruktsiooni pakkusid välja Hoshino, Kobayashi ja Suzuki [92] 2010. aastal.

Rühma- ja ringisignatuure standardib ISO. Standard ISO/IEC 20008:1 defineerib mõisted nii rühma- kui ringisignatuuridele. Standard ISO/IEC 20008:2 standardiseerib rühmasignatuure. Hetkel mustandi seisus olev standard ISO/IEC 20008:3 standardiseerib ringisignatuure. Rühma- ja ringisignatuure juurutada soovides olemasolevatele krüptoteekidele ilmselt toetuda ei saa, kui on soov luua standardile vastavaid signatuure. Küll aga võiks olla võimalik leida üsna stabiilseid avatud lähtekoodiga implementatsioone. Samuti on krüptoteekides realiseeritud „eksootilised" algebralised struktuurid ja operatsioonid, mida efektiivsemad rühma- ja ringisignatuuride konstruktsioonid kasutavad.

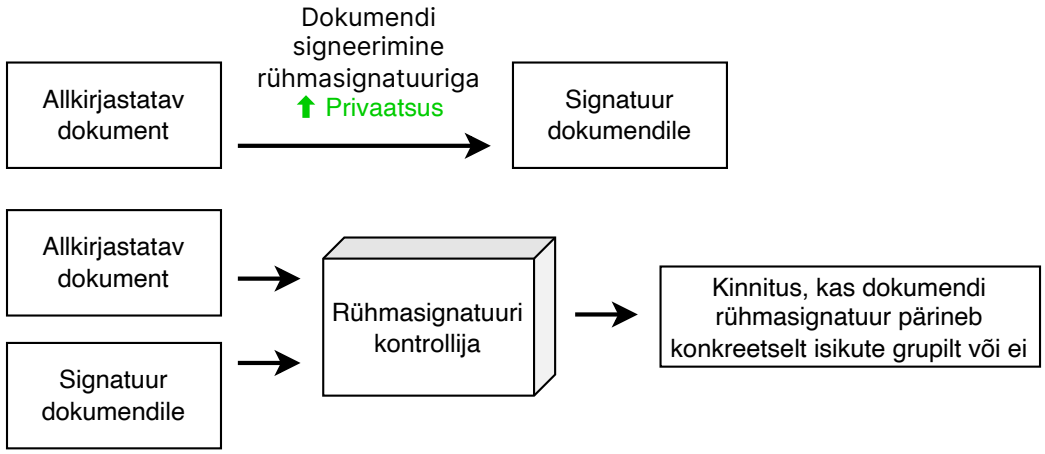
Krüptoraha Monero tehingutes kasutatakse ringi- ja rühmasignatuure selleks, et varjata ülekande sooritajat [93, 94].

**Turvagarantiid ja jääkriskid.** Rühma ja ringisignatuuride peamine terviklusomadus on sama, mis tavalistelgi signatuuridel – verifitseeruva signatuuri loomine tohib olla võimalik ainult mõnda õiget signeerimisvõtit teades. Nende peamine privaatsusomadus on, et signatuurist ei tohi olla võimalik leida selle signeerijat täpsemalt, kui see rühm, mille seas ta anonüümne on. Lisaks sellele omadusele võib aga nõuda ka tugevamaid privaatsusomadusi, näiteks mittelingitavust: kahte samale rühmavõtmele $pk^{\mathcal{G}}$ rühmasignatuuri nähes ei tohi olla arusaadav, kas need signatuurid on loonud üks ja sama isik.

**Juhised rakendajale.** Rühma- ja ringisignatuurid on komponendid suuremates infosüsteemides. Infosüsteemi arhitekt peab aru saama, kas need tehnoloogiad sobivad ehitatavasse süsteemi ja täidavad selle eesmärke.

| **Rühma- ja ringisignatuurid** | **IDENTITEET** |
|---|---|
| **Inglise keeles:** group and ring signatures | |
| **Lühidalt:** Rühmasignatuurid lubavad anda digitaalset allkirja grupi nimel. Rühmasignatuur varjab konkreetse allkirjastaja isikut. | **Arenduse keerukus:** keskmine (ringisignatuurid) / keskmine (rühmasignatuurid) |
| | **Ülalpidamise keerukus:** madal |
| | **Täpsus:** — |
| | **Privaatsusgarantii:** krüptograafiline |
| | **Tehnoloogia küpsus:** tootestatud |

**Ülevaatlik mudel:**




**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: serveri ja kliendi funktsionaalsus tuleb teostada korrektselt.
2. Jääkrisk: teostuse vea tõttu lekib konfidentsiaalseid andmeid.

**Rakendusvõimalused:**

1. Eriotstarbelised süsteemid, kus soovitakse varjata, milline konkreetne isik volitatud isikute grupist andis sõnumile digitaalse allkirja.

**Õiguspraktika:**

1. Õiguslikke hinnanguid on vähe saadaval.

**Tuntumad rakendused:**

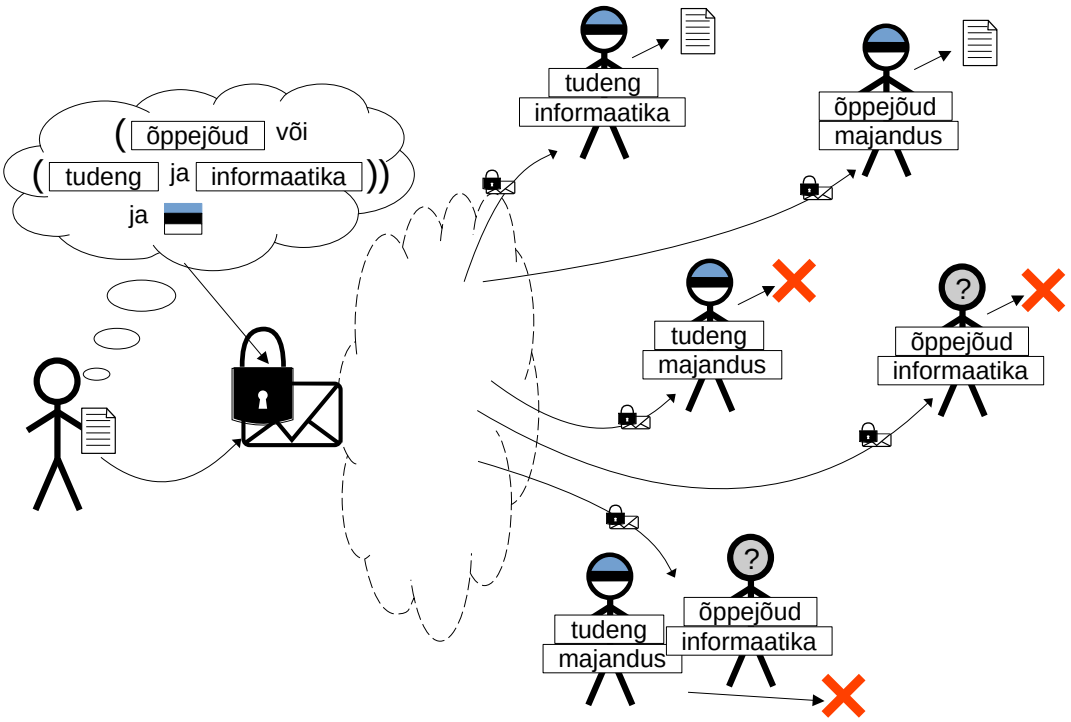
1. CryptoNote krüptovarade tehnoloogia
2. ShadowCash krüptovarade tehnoloogia
3. Monero krüptoraha

### 4.3.3 Atribuutkrüptograafia

**Lühidalt.** Atribuutkrüptograafia lubab kasutada andmete krüpteerimiseks konkreetse saaja isiku asemel tema omadusi, kaitstes nii tema identiteeti.

**Ülevaade ja rakendamine.** Atribuutkrüpteerimine on avaliku võtmega krüpteerimine, kus andmeid ei krüpteerita mitte konkreetse avaliku võtmega, millele vastab mingile konkreetsele isikule teadaolev dekrüpteerimisvõti, vaid tunnuste ehk *atribuutide* komplektiga, kus iga tunnus on midagi, mis võib andmetel ja/või isikutel olla või mitte olla. Isik saab andmed lahti krüpteerida, kui talle on antud dekrüpteerimisvõti, mille tunnused ühtuvad krüpteerimisel kasutatutega.

Joonis 3 illustreerib atribuutkrüptograafia kasutust. Sõnumi krüpteerija krüpteerib sõnumi nii, et selle saab lahti krüpteerida suvaline õppejõud või informaatika tudeng – täpsemalt siis isikud, kellele on antud vastavad tunnuste komplektid. Need, kellel on mõni nõutud tunnus puudu, sõnumit lahti krüpteerida ei saa.




**Joonis 3. Atribuutkrüptograafia illustreeriv näide**

Atribuutkrüptograafia juurutamisel on oluline roll *võtmeloomiskeskusel*; see roll on natuke sarnane sertifitseerimiskeskuse rolliga tavalises avaliku võtme infrastruktuuris. Võtmeloomiskeskus on loonud kõigile teadaoleva *avaliku alusvõtme*, mida krüpteerimisel kombineeritakse soovitud tunnuste komplektiga. Kui mingi isik soovib saada dekrüpteerimisvõtit, mis vastab teatud tunnustele, siis veenab ta võtmeloomiskeskust, et ta ise neile tunnustele vastab ja tal on õigus sellist dekrüpteerimisvõtit teada. Võtmeloomiskeskus genereerib selle dekrüpteerimisvõtme neist tunnustest ja avalikule alusvõtmele vastavast *salajasest alusvõtmest*.

Atribuutkrüptograafia leiab kasutamist kohtades, kus andmete edastamisel nende kättesaadavust tuleb kontrollida suurema hulga isikute seas. Sellisel juhul ei ole mõistlik krüpteerimise ajal täpselt määrata, millised isikud andmeid dekrüpteerida saavad, sest neid isikuid võib olla palju, nende hulk võib olla pidevas muutumises ning andmete krüpteerija ei ole suuteline otsustama, millistel isikutel on lugemisõigus.

Bezawada ja Ray [95] loetlevad mitmeid atribuutkrüpteerimise rakendusi, mis on vähemalt akadeemilist huvi ja prototüüpimist leidnud. Nende seas on

- Tasuliste telekanalite levitamine. Siin on tunnuseks õigus ühte või teist telekanalit vaadata. Kanal edastatakse krüpteerituna selle tunnusega. Klientidel on dekrüpteerimisvõtmed, millega seotud tunnused vastavad kõigile telekanalitele, mille vaatamisõiguse nad ostnud on.
- Rühmavestlused (sotsiaalvõrkudes). Rühmad vastavad tunnustele.
- Meditsiiniandmetele juurdepääsu kontrollimine. Meditsiiniprotsessides tekib palju eri tüüpi andmekirjeid, millele on lubatud ligi pääseda eri rollides isikutel: arstidel, õdedel, apteekritel, raamatupidajatel, klienditeenindajatel, audiitoritel jne. Isikute rollid on siin tunnusteks ja atribuutkrüptograafiat saab kasutada andmetele ligipääsu kontrolliks.

Bezawada ja Ray [95] nimetavad ka Zeutro LLC-d, mis üritas kasutada atribuutkrüptograafiat ettevõtte andmetele ligipääsu kontrolliks. Nad on loonud OpenABE teegi[44], mille viimane uuendus jääb kahjuks enam kui kahe aasta taha. Sarnaste kasutusjuhtude baasil üritab atribuutkrüpteerimist kommertsialiseerida NTT[45], pöörates tähelepanu ka kasutatavate krüptograafiliste konstruktsioonide postkvant-turvalisusele. ETSI on avaldanud[46] standardid atribuutkrüptograafiale ja selle kasutamisele atribuutpääsukontrollis [96] ja mõnedes teistes kasutusjuhtudes [97].

Selle jaotise esimeses lõigus kirjutasime, et dekrüpteerimine õnnestub siis, kui dekrüpteerimisvõtme tunnused *ühtuvad* krüptoteksti omadega. Mida see tähendab? Võime ette kujutada, et kusagil on komplekt tunnuseid ning mingis teises kohas on kirjeldatud poliitika, mis spetsifitseerib, millised tunnusekomplektid sobivad. Üks neist kahest (tunnusekomplekt ja poliitka) on seotud dekrüpteerimisvõtmega ja teine krüptotekstiga. Kui komplekt rahuldab poliitikat, siis dekrüpteerimine õnnestub. atribuutkrüptograafia skeeme on kaht eri tüüpi, sõltuvalt sellest, mis on seotud millega:

- Krüptotekstiga võib olla seotud tunnuste komplekt ja dekrüpteerimisvõtmega olla seotud poliitika [98]. Sellisel juhul on tunnustel tähendused, mis on kuidagi seotud krüpteeritavate andmete sisuga. Eelpool toodud rakendusnäidetest sobitub tasuliste telekanalite levitamine hästi sellise krüpteerimisskeemiga. Dekrüpteerimisvõtmega seotud poliitika kirjeldab, millise sisuga andmeid on võtmeomanikul õigus või vajadus näha.
- Dekrüpteerimisvõtmega võib olla seotud tunnuste komplekt ja krüptotekstiga olla seotud poliitika [99]. Sellisel juhul on tunnustel tähendused, mis on kuidagi seotud kasutajate rollide, õiguste ja muu säärasega. Eelpool toodud rakendusnäidetest sobitub meditsiiniandmetele juurdepääsu kontrollimine hästi sellise krüpteerimisskeemiga. Krüptotekstiga seotud poliitika kirjeldab, milliste õigustega ja/või millistes rollides isikutel on õigus andmeid näha.

Poliitika võib nõuda, et tunnuste komplekt peab sisaldama mingit konkreetset tunnust, või mitut konkreetset tunnust, või vähemalt ühte mitmest võimalikust tunnusest, või ka mingit keerulisemat kombinatsiooni tunnustest. Erinevad krüptograafilised konstruktsioonid atribuutkrüpteerimiseks toetavad eri väljendusvõimsusega poliitikaid [100], samuti sõltub nende jõudlus eri viisil parameetrite (tunnuste ja kasutajate arv, tunnuste arv tunnuste komplektis, poliitika keerukus) suurusest, mis atribuutkrüpteerimise juurutamisel olulised võivad olla. Üks huvitav aspekt, milles eri konstruktsioonid erinevad, on tugi *negatiivsetele* tunnustele: kas (näiteks) krüptoteksti poliitika saab spetsifitseerida, et dekrüptimine on võimalik ainult siis, kui dekrüpteerimisvõtmega mingit teatud tunnust seotud *ei ole*? Negatiivsed tunnused iseenesest ei anna poliitikate väljendusvõimsusele midagi juurde: samasugune tulemus on võimalik saavutada *monotoonsete poliitikatega* (s.t. kui mingi tunnusekomplekt rahuldab poliitikat, siis mõni teine komplekt, mis seda sisaldab, rahuldab seda poliitikat samuti), kui me lisaks igale tunnusele $A$ toome sisse tunnuse $\overline{A}$, mille tähendus on tunnuse $A$ puudumine; iga tunnusekomplekt peab sisaldama täpselt ühte tunnustest $A$ ja $\overline{A}$. Küll aga võib selline teisendus juurutust ebaefektiivsemaks muuta, sest et nüüd on kõigis kasutusel

[44] https://github.com/zeutro/openabe (viimati külastatud 01.03.2023).

[45] NTT is developing attribute-based encryption (ABE) to prevent quantum attacks https://venturebeat.com/security/ntt-is-developing-attribute-based-encryption-abe-to-prevent-quantum-attacks (viimati külastatud 01.03.2023).

[46] ETSI releases cryptographic standards for secure access control https://www.etsi.org/newsroom/press-releases/1328-2018-08-press-etsi-releases-cryptographic-standards-for-secure-access-control (viimati külastatud 01.03.2023).

olevates tunnusekomplektides palju tunnuseid.

**Ajaloost.** Atribuutkrüpteerimise pakkusid välja Sahai ja Waters [101] 2005. aastal. Nad pakkusid selle välja kui üldistuse *identiteedipõhisele krüptograafiale* [102, 103], kus isiku avalikuks võtmeks oli tema nimi ning võtmeloomiskeskus oli võimeline igale isikule genereerima tema privaatvõtme.

**Turvagarantiid ja jääkriskid.** Tegemist on krüpteerimisprimitiiviga, nii et peamised turvaomadused, mida me atribuutkrüpteerimiselt ootame, on samad, mida avaliku võtme krüptosüsteemidelt: võimatus ilma dekrüpteerimsvõtit teadmata leida teatud krüptotekstist avateksti, isegi kui meil on juurdepääs funktsionaalsusele, mis dekrüpteerib teisi krüptotekste (turvalisus valitud krüptotekstiga ründe vastu) ning genereerib dekrüpteerimisvõtmeid, millega seotud tunnusekomplekt või poliitika ei ühtu krüptoteksti poliitika või tunnusekomplektiga.

Primitiivile spetsiifilisem turvaomadus on võimatus kombineerida erinevaid dekrüpteerimisvõtmeid. Näiteks skeemi korral, kus dekrüpteerimisvõtmetega on seotud tunnusekomplektid, ei tohi olla võimalik leida krüptotekstile vastavat avateksti, kui ükski meie käsutuses olev võti ei ole seotud tunnusekomplektiga, mis krüptotekstiga seotud poliitikat rahuldab; seda isegi siis, kui seda poliitikat rahuldab meie käsutuses olevate võtmete tunnusekomplektide *ühend*.

**Juhised rakendajale.** Atribuutkrüpteerimise tehnoloogiate pakkujaid pigem ei leidu, küll leidub rohkem või vähem küpseid tarkvarateekide realisatsioone. Kui seda tehnoloogiat aga juurutama hakata, siis tuleb kindlasti läbi mõtelda, kus asub võtmeloomiskeskus. Selle süsteemikomponendiga on seotud väga tugevad usaldusnõuded. Vajadusel võib kaaluda mõne usaldustnõudeid vähendava tehnoloogia (näiteks turvaliste ühisarvutuste) kasutamist võtmeloomiskeskuse realisatsioonil. Võtmeloomiskeskuse asukoha valik võib olla veelgi komplitseeritum juhul, kui süsteemis on tunnuseid, mis loogiliselt kuuluvad erinevatesse domeenidesse, nii et otsustaja, kas kasutajal on või ei ole teatud tunnust, on eri domeenidesse kuuluvate tunnuste korral erinev. Sel juhul võivad need otsustajad kasutajale väljastada tõendeid tema tunnuste kohta; võtmeloomiskeskus loeb neid tõendeid ja genereerib kõigi vajalike tunnustega dekrüpteerimisvõtme. Alternatiivina on võimalik igas domeenis kasutada eri võtmeloomiskeskust [104].

Atribuutkrüpteerimise (sarnaselt identiteedipõhise krüpteerimisega) üks kitsaskohti on võtmete tühistamise keerukus. Kui mingi dekrüpteerimisvõti on lekkinud, siis kuidas vältida selle edasist kasutust, arvestades, et me ilmselt ikkagi soovime luua krüptotekste, mille poliitika / tunnusekomplekt ühtuks selle võtme tunnusekomplekti / poliitikaga. Häid lahendusi siin ei ole, üks võimalus on kasutada tunnuseid, mis vastavad (üsna lühikestele) ajavahemikele. Iga sellise ajavahemiku alguses peavad kasutajad endale uued võtmed küsima, kus tunnusekomplekt sisaldab sellele ajavahemikule vastavat tunnust või poliitika nõuab selle tunnuse olemasolu. Isikule, kelle dekrüpteerimisvõti on lekkinud, võtmekeskus uut võtit ei genereeri. Kõik selle ajavahemiku jooksul loodavad krüptotekstid peavad sedasama tunnust nõudma või sisaldama, seega ei saa lekkinud võtmega isik neid enam dekrüpteerida. Teine võimalus on kasutada konstruktsiooni, mis toetab negatiivseid tunnuseid, luua üks tunnus iga kasutaja kohta (ja lisada see sellele kasutajale väljastatava dekrüpteerimisvõtme tunnusekomplekti) ning mingi kasutaja võtme lekkimisel hakata edaspidi loodavate krüptotekstidega seotud poliitikates nõudma, et sellele kasutajale vastav tunnus ei tohi tunnusekomplekti kuuluda.

**Signeerimine.** Maji, Prabhakaran ja Rosulek on välja on pakutud ka *atribuutsignatuurid* [105], kus võtmeloomiskeskus väljastab kasutajatele signeerimisvõtmeid, mis on seotud selle kasutaja kohta käivate tunnuste komplektiga. Kui kasutaja loob signatuuri, siis on tal võimalik valida, millisele poliitikaga see signatuuri seotakse; valitav poliitika peab olema aga selline, et kasutaja enda tunnuste komplekt seda poliitikat rahuldaks. Signatuuri verifitseerimise üheks argumendiks (lisaks signeeritud sõnumile, signatuurile ja avalikule alusvõtmele) on seesama poliitika. Atribuutsignatuuride oluline privaatsusomadus on signatuuride omavahel seostamatus – mitut signatuuri nähes ei tohi olla võimalik aru saada, kas need on loodud sama kasutaja poolt. Atribuutsignatuurid omavad teatavat sarnasust ringisignatuuridega – signeerimisel teatud poliitika valimise asemel võiks moodustada rühma kõigist neist isikutest, kellega seotud tunnusekomplektid valitavat poliitikat rahuldavad. Erinevus tuleb aga sellest, et atribuutsignatuure kasutades ei pea signeerija kõiki neid isikuid ja nende avalikke võtmeid teadma.

## Atribuutkrüptograafia

**IDENTITEET**

**Inglise keeles:** attribute based cryptography

**Lühidalt:** Atribuutkrüptograafia lubab kasutada andmete krüpteerimiseks konkreetse saaja isiku asemel tema omadusi, kaitstes nii tema identiteeti.

**Arenduse keerukus:** kõrge

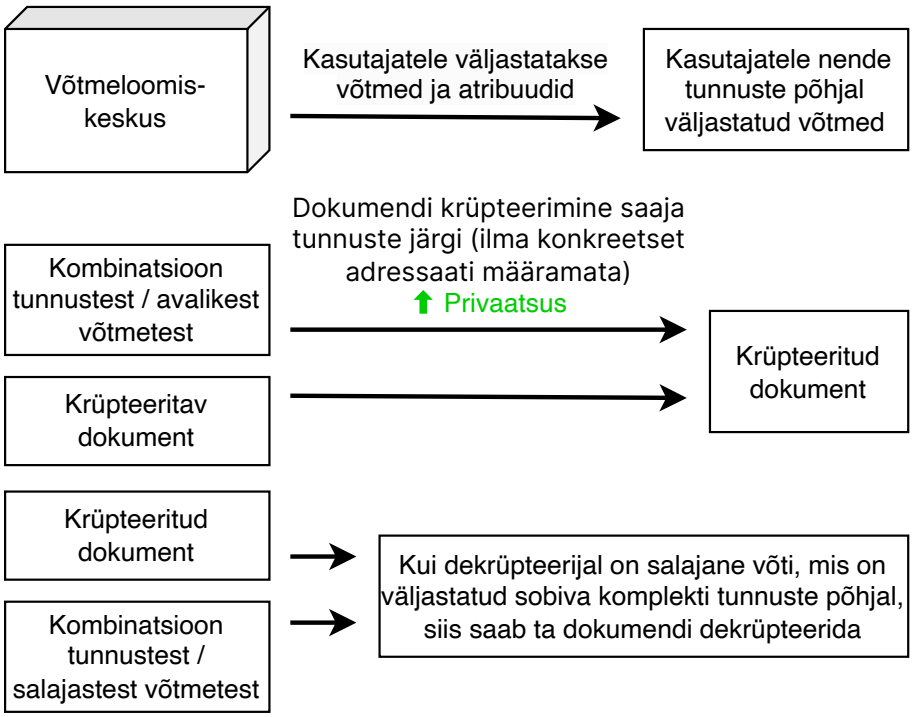
**Ülalpidamise keerukus:** madal

**Täpsus:** täpne

**Privaatsusgarantii:** krüptograafiline

**Tehnoloogia küpsus:** madal

**Ülevaatlik mudel:**




**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: kasutajatele dekrüpteerimisvõtmeid väljastav võtmeloomiskeskus peab korrektselt töötama.
2. Turvaeeldus: tehnoloogia peab arenema nii kaugele, et lahendatud saavad rakenduslikud küsimused nagu võtmete tühistamine.
3. Jääkrisk: tehnoloogia või teostuse vea tõttu lekib konfidentsiaalseid andmeid.

**Rakendusvõimalused:**

1. Andmete turvaline jagamine olukorras, kus rollipõhist juurdepääsukontrolli ei saa usaldada.

**Õiguspraktika:**

1. Tehnoloogial pole juurutusi, mis oleks nõudnud õiguslikku analüüsi.

**Tuntumad rakendused:**

1. Tehnoloogia on arenduses ning tööstuslikke rakendusi ei ole veel teada.

### 4.3.4 Nullteadmustõestused

**Lühidalt.** Nullteadmustõestusega saab privaatsete andmete teadja tõestada kellelegi teisele, et tema andmetel on mingi omadus, ilma neid andmeid avaldamata.

**Ülevaade ja rakendamine.** Nullteadmustõestused on protokollid, mis lubavad ühel osapoolel – Tõestajal – teisele osapoolele – Kontrollijale – tõestada ilma oma käsutuses olevaid andmeid näitamata, et andmed on mingis vahemikus, vastavad mingile standardile või neist saab arvutada mingi avaliku tulemuse. Kontrollija saab teada ainult selle, kas tõestus läks läbi või ei. Kui Tõestaja käsutuses olevad andmed ei rahulda neid omadusi, mida ta tõestada püüab, siis ei suuda ta Kontrollijat veenda isegi juhul, kui ta protokollist kõrvale kaldub.

Nullteadmustõestuse juurutamisel on vaja spetsifitseerida väide, milles Tõestaja Kontrollijat veenab. See väide on esitatav kui programm, mis tagastab tõeväärtuse. Ta võtab kaks argumenti: andmed, mis on teada nii Tõestajale kui ka kontrollijale, ja andmed, mis on teada ainult Tõestajale. Näiteks graafi tippude värvimise ülesande juures on mõlemale teada graafi struktuur, aga ainult Tõestajale teada tippude värvid. Programm käib läbi graafi kõik servad ja kontrollib, et otstipud oleksid eri värvi.

Väidete, mida me tegelikult tõestada soovime nii krüptograafiliste konstruktsioonide juures kui ka talitlusmudelites, tõlkimine graafi tippude värvitavuseks on teoreetiliselt küll efektiivne, kuid praktikas kaugelt liiga keeruline. Juurutajale mugavam on spetsifitseerida väide kas loogika- või aritmeetilise lülitusena; ka selliselt esitatud väidete jaoks leiduvad nullteadmustõestused, mis põhinevad diskreetsete logaritmide omavaheliste suhete demonstreerimisel [106].

Kahjuks on selle protokolli praktiline efektiivsus suhteliselt väike, sest ta peab iga lülituses oleva värava jaoks arvutama mitu modulaareksponenti. Kui aga väide on aritmeetilise lülitusena spetsifitseeritud, siis võime vaadata nullteadmustõestusi kui turvalise ühisarvutuse eriliiki, kus meil on täpselt kaks osapoolt, kellest üks teab arvutuse kõiki sisendeid, aga võib protokolli mitte täita. Peale piisavalt efektiivsete turvalise ühisarvutuse protokollide väljapakkumist, mis on turvalised ka protokollist kõrvale kalduvate osapoolte korral, hakati neid protokolle kohandama ka nullteadmustõestustele. Osad efektiivseimad nullteadmustõestuseprotokollid [107, 108] on konstrueeritud sel põhimõttel.

Teine liik nullteadmustõestuste konstruktsioone põhineb *verifitseeritavatel arvutustel*. Need on protokollid, mille osapooled ja turvanõuded on põhimõtteliselt samad, mis nullteadmustõestustel, aga Tõestaja privaatsust ei garanteerita. Selliste protokollide eesmärgiks on vähendada protokolli sõnumite pikkust ja Kontrollija tööd; tuleb välja, et sõnumite pikkus on võimalik muuta sõltumatuks arvutuse keerukusest [109] ning ka Kontrollija tööaeg saab olla oluliselt väiksem kui arvutuse enda teostamiseks vajalik [109, 110]. Verifitseeritavate arvutuste muutmine Tõestaja privaatsust säilitavaks on tüüpiliselt lihtne võrreldes verifitseerimisprotokolli enda kontseptuaalse keerukusega.

**Nullteadmustõestuse protokollide võimalikud kujud.** Nullteadmustõestused on protokollid kahe osapoole – Tõestaja ja Kontrollija – vahel. Need protokollid tüüpiliselt nõuavad paljude teadete vahetamist nende kahe osapoole vahel; teadete arv sõltub tõestatavat väidet kodeeriva loogika- või aritmeetilise lülituse kujust. Protokolli eduka lõpetamise korral on Kontrollija väite kehtivuses veendunud. Ta ei saa seda veendumust aga edasi anda – protokolliteadete näitamine kolmandale osapoolele seda osapoolt enam ei veena, sest Kontrollija oleks võimeline eristamatu teadetejärjendi looma ilma Tõestajaga suhtlematagi (see ongi põhimõtteliselt klassikaline „nullteadmuse“ definitsioon).

Hoopis teistsugused on *mitteinteraktiivsed nullteadmustõestused*. Siin saadab (või publitseerib) Tõestaja üheainsa teate, Kontrollija loeb seda, teeb mingid kontrollivad arvutused ja saab veendud, et väide kehtib. Mitteinteraktiivsetel tõestustel on seega mitu head külge: veenda on võimalik mitut Kontrollijat ja Kontrollija tehtavad arvutused võivad ka väiksemad olla. Halvaks küljeks on Tõestaja suurem arvutusmaht.

Leidub viis, nn. Fiat-Shamiri heuristika [111], mis teatud klassi interaktiivseid protokolle muudavad mitteinteraktiivseks. Tõestaja ja Kontrollija töömaht seejuures oluliselt ei muutu. See heuristika pole aga rakendatav efektiivseimatele interaktiivsetele protokollidele, mida eespool mainisime. Lühemate tõestuste ja Kontrollija väiksemate arvutustega (mitteinteraktiivsed) protokollid on loodud teisiti. Siin peab nii Tõestajale kui ka Kontrollijale kättesaadav olema sõne, millel on kindel struktuur [112]. See struktuur võib sõltuda konkreetsest väitest, mida tõestatakse (kuid on sõltumatu argumentidest, mis sellele väitele kui programmile ette antakse). Mõne teise protokollikonstruktsiooni korral on see struktuur sõltumatu konk-

reetsest väitest, aga sõne peab olema piisavalt pikk (sõltub väidet kodeeriva lülituse suurusest). Igal juhul peab sõnel olema kindel struktuur; iga juhuslik bitijada ei sobi. Vale struktuuriga sõne rikub ära protokolli turvaomadused. Samuti on turvaomadused rikutud, kui lekivad teatud juhuslikud väärtused, mille alusel selle sõne struktuur tekitatakse.

Näeme, et selle struktuurse sõne peab genereerima mingi usaldatud osapool. Üks võimalus selline osapool leida on leida palju enam-vähem usaldusväärseid isikuid ning lasta neil see struktuurne sõne genereerida turvalist ühisarvutust kasutades. Selline arvutus võib olla kallis, aga neid on tehtud[47]. Osadel nullteadmustõestusprotokollidel on võimalik struktuurne sõne genereerida turvalise ühisarvutuse (ptk 4.2.10 abil niimoodi, et kõik arvutuse osapooled ei pea üksteisega teateid vahetama. Selle asemel käib suhtlus niimoodi, et esimene arvutuse osapool loob esialgse sõne, peale mida saadetakse see läbi kõigi osapoolte, kes seda sõnet uuendavad. Kui vähemalt üks osapooltest on aus, siis on saadav struktuurne sõne sobilik protokollis kasutamiseks.

Lisaks interaktiivsetele, ühe võimaliku Kontrollijaga protokollidele ja mitteinteraktiivsetele, suvalise Kontrollijaga protokollidele leidub ka kombinatsioon nendest: mitteinteraktiivsed protokollid, mille loodavad tõestused veenavad ära üheainsa võimaliku kontrollija. Sellised protokollid on ühtpidi paindlikumad kui interaktiivsed, teistpidi efektiivsemad kui suvalise Kontrollijaga.

Esimese krüptovaluutana kasutas nullteadmustõestusi Zerocoin [113], mis pakkus vaid osalist anonüümsust. Nullteadmustõestused võimaldavad krüptorahade Zcash ja Monero tehinguid valideerida ilma nendega seotud aadresse ja ülekantavat summat avaldamata [114]. Monero anonüümsete tehingute arv on tuhandetes ja nende kaudu on kantud üle miljardite eurode väärtuses vara [115].

Krüptorahad, millega on võimalik anonüümselt tehinguid teha, on leidnud kasutust ka ebaseaduslikeks tegevusteks nagu rahapesu, lunavara või kaevekaaperdus. Hoolimata sellise tegevuse väikesest mahust on mitmed valitsused hakanud nende kasutust reguleerima, piirama või keelama [116].

**Ajaloost.** Nullteadmustõestuste väljapakkujateks loetakse Goldwasserit, Micalit ja Rackoffi [117] (1985), kes selle eest 1993. aastal jagasid Gödeli auhinda. Nad näitasid nullteadmustõestuse leidumist ühele konkreetsele väitetüübile, mis leiab kasutust krüptograafilistes konstruktsioonides. Järgmisel aastal näitasid Goldreich, Micali ja Wigderson [118, 119], et nullteadmusega on võimalik tõestada, et mingi graafi tipud on värvitavad kolme värviga, nii et iga serva otstipud on eri värvi. See on üldine tulemus, sest et graafi tippude kolme värviga värvitavusele on *teoreetiliselt* vähese vaevaga taandatavad kõik väited, mille tõestust on arvutuslikult lihtne kontrollida.

Nullteadmustõestused tulid kõigepealt kasutusele krüptograafiliste konstruktsioonide osana, kus tõestati teatud kujul väiteid, mille jaoks osati vajalikke protokolle koostada. Tüüpiliselt olid need väited teatud diskreetsete logaritmide teadmise, võrdsuse, või muude algberaliste omaduste kohta. Aastal 1989 välja pakutud Schnorri signatuurid ja identifitseerimisskeem [120] olid üks esimesi konstruktsioone, mille põhiosa oli nullteadmustõestus diskreetse logaritmi teadmise kohta. Teine varane näide nullteadmustõestuste kasutamisest on teadetetahvlil põhinevad e-hääletusprotokollid [121], kus hääletaja kirjutab tahvlile oma krüpteeritud hääle ning annab tõestuse, et hääl on korrektsel kujul (kombinatsioon teatud diskreetsete logaritmide teadmisest ja võrdsusest). Nullteadmustõestused on ka valitud krüptotekstiga rünnete suhtes turvaliste avaliku võtme krüpteerimisskeemide osa [122]. Selliste nullteadmustõestuste üldistamine viis nn. Groth-Sahai tõestusteni [123], millele tuginevad osad efektiivseimad konstruktsioonid tänapäeval.

**Juhised rakendajale.** Soovides nullteadmustõestusi mingis infosüsteemis kasutada, tuleks kõigepealt läbi mõtelda, kas sealne kasutusjuht on tegelikult nullteadmustõestuste kasutusjuht – kas seal on Tõestaja ja Kontrollija, kus Tõestaja teab kõike ja Kontrollija eesmärk on üksnes kindlaks teha, et Tõestaja ei valeta? Kui tundub, et Kontrollijal on ka mingid andmed, mida Tõestaja ei tea, siis on tegemist üldisema ülesandega, mis on ilmselt turvalise ühisarvutuse kasutusjuht.

Kui on selge, et leidub osapool „Tõestaja“, kes teab kõike, siis tuleks järgmisena läbi mõtelda, mis on see argument tõestatavat väidet esitavale programmile, mida teab ka Kontrollija. Teisisõnu, mis on see reaaleluline, (suhteliselt) avalik ankur, mille suhtes me näitame, et tõestatav väide kehtib? Selleks ankruks võib olla näiteks mingisugune avalik võti, mille omanik on süsteemis hästi äratuntav. Selleks ankruks võib

[47] The crazy security behind the birth of Zcash, the inside story https://spectrum.ieee.org/the-crazy-security-behind-the-birth-of-zcash (viimati külastatud 01.03.2023).

olla mingi bitijada, mille Kontrollija leiab moel, mis ei sõltu täielikult Tõestajast. Kui sellist ankrut ei leidu, siis ei ole meil tegelikult võimalik Tõestaja esitatavat väidet kontrollida.

Peale neile küsimustele vastamist saab läbi mõtelda juurutuse detailid. Kui palju on Tõestajaid ja Kontrollijaid, millise kujuga nullteadmustõestuse protokolli saab kasutada? Kuidas seada üles usaldatud osapool, kui seda peaks struktuurse sõne genereerimiseks vaja olema?

Avatud lähtekoodiga nullteadmustõestuste realisatsioone on olemas mitmeid, samuti leiab neid osana suurematest süsteemidest. Oluline rakendus, mis nende realisatsioonide küpsemist veab, on privaatsust säilitav krüptoraha plokiahelal. Selles valdkonnas leidub idufirmasid ja standardimisinitsiatiive[48].

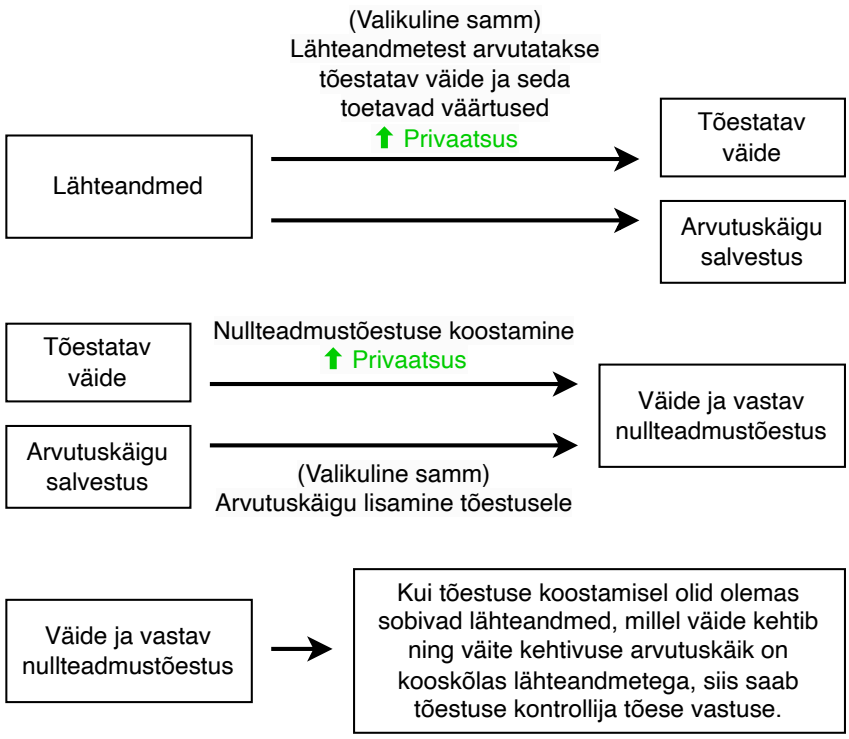
**Õiguslikud aspektid.** Nullteadmustõestuste tehnoloogial on potentsiaali parandada andmete privaatsust ja turvalisust paljudel kasutusjuhtudel ning toetada regulatiivsete piirangute täitmist[49]. Siiski ei ole tehnoloogia enda kohta õiguslikke seisukohti saadaval.

[48] ZKProof Standards https://zkproof.org (viimati külastatud 01.03.2023).

[49] Zero Knowledge Proof: how to maintain privacy in a data-based world https://www.bbva.com/en/zero-knowledge-proof-how-to-maintain-privacy-in-a-data-based-world (viimati külastatud 25.01.2023).

| **Nullteadmustõestused** | **IDENTITEET** |
|---|---|
| **Inglise keeles:** zero knowledge proofs | |
| **Lühidalt:** Nullteadmustõestusega saab privaatsete andmete teadja tõestada kellelegi teisele, et tema andmetel on mingi omadus, ilma neid andmeid avaldamata. | **Arenduse keerukus:** kõrge |
| | **Ülalpidamise keerukus:** keskmine |
| | **Täpsus:** täpne |
| | **Privaatsusgarantii:** matemaatiliselt tõestatav |
| | **Tehnoloogia küpsus:** keskmine |

**Ülevaatlik mudel:**




| **Turvaeeldused ja jääkriskid:** | **Rakendusvõimalused:** |
|---|---|
| 1. Turvaeeldus: tõestatav väide tuleb õigesti kirja panna (s.t. ära programmeerida), kasutades sobivat nullteadmuse tehnoloogiat.<br>2. Turvaeeldus: kui kasutatav protokoll vajab eeltöötlust usaldatud kolmanda osapoole poolt, siis tuleb see korrektselt teostada.<br>3. Jääkrisk: teostuse vea tõttu lekib konfidentsiaalseid andmeid. | 1. Tõendite esitamine (nt kukrutes või isikutunnistustes) |
| **Õiguspraktika:**<br>1. Õiguslikus mõistes täiendavad kaitsemeetmed. | **Tuntumad rakendused:**<br>1. Internetihääletamise süsteemid<br>2. Zcash krüptoraha |

# Nullteadmustõestused ning ringi- ja rühmasignatuurid krüptorahades

**Lühidalt:** Krüptorahad Zcash ja Monero kasutavad nullteadmustõestuseid, et valideerida tehinguid ilma nendega seotud aadresse ja ülekantavat summat avaldamata. Monero tehingutes kasutatakse ka ringi- ja rühmasignatuure, mis võimaldavad varjata ülekande sooritajat.

**Teostamise aasta:** Alates 2013

**Riik:** Globaalne ja detsentraliseeritud

**Omanik:** detsentraliseeritud

**Teostaja:** detsentraliseeritud

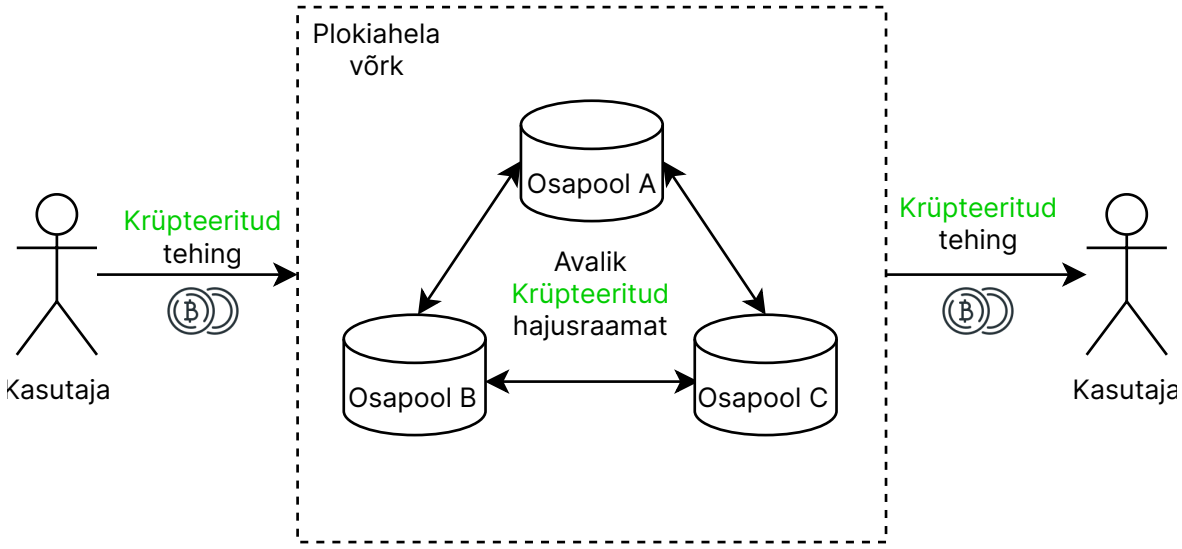
**Süsteemi küpsus:** püsiv juurutus

**Privaatsuskaitse tehnoloogiad:**

1. nullteadmustõestused
2. otspunktkrüpteerimine
3. pimesignatuurid
4. ringi- ja rühmasignatuurid

**Sobivad kasutusjuhtumid:**
hajutatud varade platvorm

**Ülevaatlik mudel:**




Kasutajad saadavad krüpteeritud tehinguid plokiahela võrgule valideerimiseks ilma nendega seotud andmeid avaldamata.

⬆Privaatsus

⬇Auditeeritavus

**Märkimisväärsed omadused:**

1. Krüptorahade tehingute üle arve pidamisel kasutatakse hajusraamatuid, mis on tihti täiesti avalikud ja kasutavad privaatsuse tagamiseks vaid kasutajate pseudonüüme. Privaatsuskaitse tehnoloogiad võimaldavad tagada süsteemi kasutajate anonüümsuse ja maksete korrektsuse.
2. Zcash ja Monero on tehnoloogiat edasi arendanud ja juurutanud.
3. Privaatsuskaitse tehnoloogiaid nähakse krüptorahadega tehtud maksete juures kui sammu privaatsuse ja anonüümsuse suunas ning seetõttu on vastavaid tehnoloogiaid kasutavad krüptorahad osutunud väga populaarseks. Igapäevaste anonüümsete tehingute arv on tuhandetes ja nende kaudu kantakse iga päev üle miljardeid dollareid.
4. Krüptorahad, millega on võimalik anonüümselt tehinguid teha, on mõningal määral leidnud kasutust ebaseaduslikeks tegevusteks nagu rahapesu, lunavara või arvutusressursside kaaperdamine.

## 4.4 Anonüümne side ja tehingud

### 4.4.1 Turvaline vestlus

**Lühidalt.** Turvaline vestlus lubab kahel osapoolel omavahel privaatselt ja segamatult suhelda.

**Ülevaade ja rakendamine.** Turvalise vestluse tehnoloogiad kasutavad krüptograafilisi vahendeid, et tagada sõnumite konfidentsiaalsus ja terviklus. Tüüpiliselt tähendab see, et sõnumi saatja peab sõnumi krüpteerima, et muuta see kõigile teistele loetamatuks. Sõnumi saaja on ainus, kes suudab sõnumi dekrüpteerida ning selle autentsuses veenduda. Kõik see eeldab nii sümmeetriliste kui ka avaliku võtme krüptosüsteemide kasutamist, ms põhinevad salajastel võtmetel või võtmepaaridel.

Tuntumad turvalist suhtluskanalit pakkuvad tehnoloogiad ja rakendused on turvalist võrguühendust pakkuv TLS, turvalist meilivahetust pakkuv PGP, otspunktkrüpteerimist võimaldavad WhatsApp, Signal [124, 125] ja Proton ning mitmed video- ja häälkõne rakendused. Neist Signal on kaasatud Elektronhõlve Fondi salajase jälgimise vastase kaitse juhistesse[50]ja lisaks sai rakendus samalt fondilt turvalise sõnumside rakenduste tulemuskaardil ideaalse tulemuse [126].

Mitmed riigid, valitsused ja asutused on Signali kasutamist keelanud või piiranud, tuues probleemidena välja selle kasutuse ebaseaduslikeks ja ohtlikeks tegevusteks ning kuritegevuslikes organisatsioonides [127].

**Ajaloost.** Vajadus sõnumisaladuse järele oli olemas kaugelt enne arvutivõrgu tekkimist ning seepärast teame esimesi šifreid juba antiikajast. Tänapäeval on turvaline vestlus osa elektroonilistest suhtlusrakendustest, pakkudes kaitset küberrünnakute ning seiresüsteemide vastu [128]. Isiklikud, ärilised ja riiklikud vajadused on tinginud üha võimekamate ja turvalisemate suhtlusprotokollide arenduse, et kaitsta erinevat liiki tundlike andmeid, nagu ärisaladus, terviseandmed, juriidilised või ka isikuandmed. Esimene avalikult kasutatav võrgukihi turvet pakkuv lahendus oli 1995. aastal avalikustatud *Secure Sockets Layer* (SSL) versioon 2.0. SSL ja tema järeltulija *Transport Layer Security* (TLS) eeldasid krüptograafia kasutamise tõttu võimekamat riistvara ning teenuseandjalt ka serveri sertifikaatide eest tasumist. Seetõttu kasutasid neid tehnoloogiaid pikka aega vaid väga IT-võimekad teenuseandjad (nt pangad ja tehnoloogiaettevõtted). TLS kasutus kasvas aga hüppeliselt peale Snowdeni lekkeid 2013. aastal kui üldsus sai aru, et nende võrguühendusi pealt kuulatakse[51].

Maailm on liikunud üha rohkem inimesekesksemaks, kus turvaline ühenduskanal kasutaja seadme ja teenusepakkuja vahel ei ole enam piisav, et tagada nõutav privaatsuse tase. Suhtlusrakendustelt nõutakse otspunktkrüpteerimist (ingl k *end-to-end encryption*, E2EE), kus katkematu turvaline ühenduskanal on loodud kahe lõppseadme vahel. Suhtlus võib küll toimida teenuseandja serveri vahendusel, kuid see ei saa sõnumeid lugeda ega muuta.

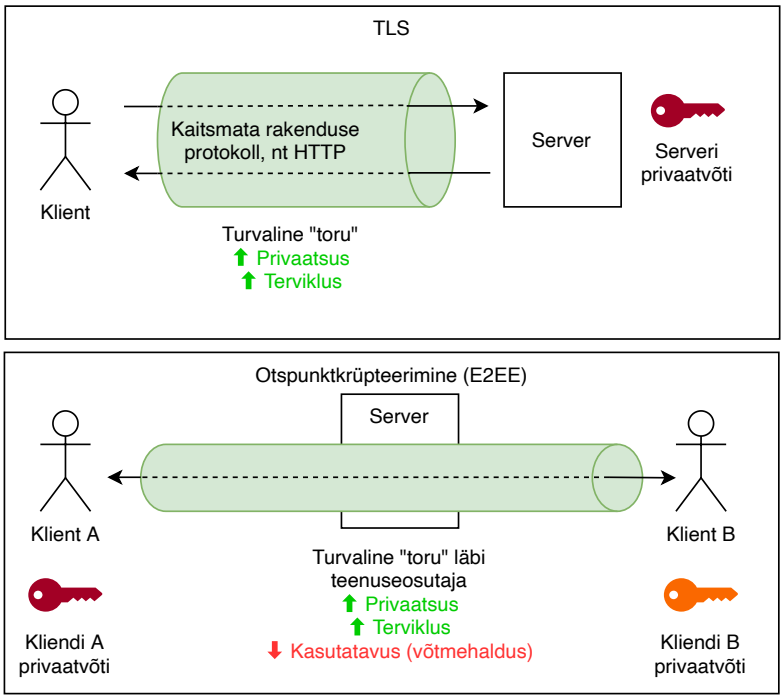
**Turvagarantiid ja jääkriskid.** Turvalise vestluse süsteemid on loodud tagama suhtuse terviklust ja konfidentsiaalsust. Seeläbi saavutatakse privaatne vestlus, kus volitamata osapooled ei saa sõnumeid lugeda ega muuta. Mõned süsteemid võimaldavad vestluse osapoolte isikut ka verifitseerida, pakkudes sõnumite autentsust ja salgamatust. Turvalise vestluse süsteem eeldab, et vestluse otspunkt ise on turvaline ning seetõttu ei paku kaitset näiteks otspunktis oleva kahjurvara või seal aset leidva teenusetõkestusründe vastu. Samuti on süsteem haavatav kui salajased võtmed on lekkinud. Otspunktkrüpteeritud süsteemide puhul tasub tähele panna, et kõrge privaatsustaseme tagamiseks peavad lõppkasutajad ise oma krüptovõtmeid haldama.

[50] Communicating with Others https://ssd.eff.org/module/communicating-others (viimati külastatud 02.03.2023).

[51] Matthew Green. Looking back at the Snowden revelations, 2019. https://blog.cryptographyengineering.com/2019/09/24/looking-back-at-the-snowden-revelations (viimati külastatud 01.03.2023).

| **Turvaline sõnumivahetus** | **SIDE** |
|---|---|
| **Inglise keeles:** secure messaging | |
| **Lühidalt:** Turvaline vestlus lubab kahel osapoolel omavahel privaatselt ja segamatult suhelda. | **Arenduse keerukus:** madal |
| | **Ülalpidamise keerukus:** madal |
| | **Täpsus:** *ei rakendu* |
| | **Privaatsusgarantii:** krüptograafiliselt tõestatav |
| | **Tehnoloogia küpsus:** kõrge |

**Ülevaatlik mudel:**




**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: otspunktid ei tohi olla kompromiteeritud.
2. Turvaeeldus: otspunktkrüpteerimise puhul peavad kasutajad või nende seadmed ise võtmehaldusega tegelema.
3. Jääkrisk: andmed lekivad otspunktides.

**Rakendusvõimalused:**

1. Turvaline side arvutivõrkudes

**Õiguspraktika:**

1. Turvalise sõnumivahetuse tehnoloogiad on õiguslikkus mõistes täiendavad kaitsemeetmed.
2. Mõnedes piirkondades kohustatakse ka turvalise sõnumivahetuse tehnoloogia pakkujaid looma tehnilisi meetmeid õiguskaitseorganite nõudmisel sõnumisaladuse riivamiseks.

**Tuntumad rakendused:**

1. Turvamata protokollide turvaliseks muutmine TLS abil (näiteks HTTPS, IMAPS, POP3S, jne)
2. Turvaline e-postivahetus (näiteks PGP)
3. Turvaline otsesuhtlus (näiteks Signal, WhatsApp, iMessage)
4. Video- ja häälkõne rakendused (näiteks Zoom)

## Signal

**Lühidalt:** Signal on suhtlusrakendus, mille kaudu saab saata otspunktkrüpteeritult kiirmeile, teha kõnehelistusi ja videokõnesid. Rakenduse fookus on privaatsusel ja turvalisusel ning selle klienditarkvara on vabavaraline ja avatud lähtekoodiga.

**Teostamise aasta:** Alates 2013

**Riik:** Ameerika Ühendriigid

**Omanik:** Signal Foundation

**Teostaja:** Algne arendaja Whisper Systems, täna arendab Signal Foundation

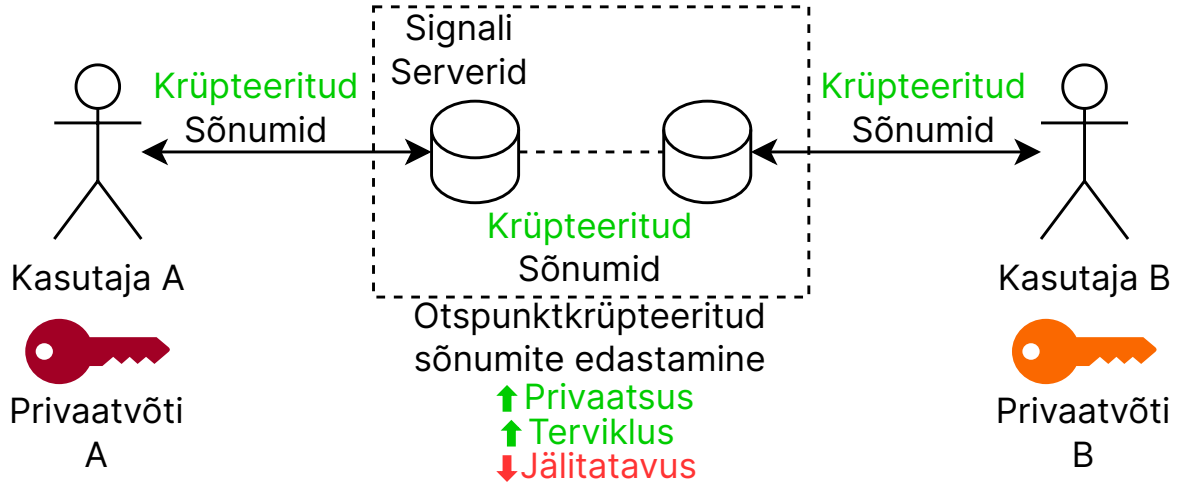
**Süsteemi küpsus:** püsiv juurutus

**Privaatsuskaitse tehnoloogiad:**

1. otspunktkrüpteerimine
2. usaldatavad käivituskeskkonnad
3. nullteadmus

**Sobivad kasutusjuhtumid:**
otspunktkrüpteeritud suhtlus

**Ülevaatlik mudel:**




**Märkimisväärsed omadused:**

1. Rakenduse igakuine aktiivsete kasutajate arv on ligikaudu 50 miljonit.
2. Signal on ekspertide poolt kiidetud ning võitnud mitu granti ja auhinda. Elektronhõlve Fond lisas rakenduse salajälgimise vastase kaitse juhistesse, lisaks sai rakendus samalt fondilt turvalise sõnumside rakenduste tulemuskaardil ideaalse tulemuse.
3. Mitmed riigid, valitsused ja asutused on Signali kasutamist keelanud või piiranud, tuues probleemidena välja raskuse selle pealt kuulamisel ning kasutuse ebaseaduslikeks ja ohtlikeks tegevusteks ning kuritegevuslikes organisatsioonides.

### 4.4.2 Mikservõrgud

**Lühidalt.** Mikservõrk võtab mitu privaatset sõnumit ja segab nad omavahel ära nii, et vaatleja ei saa aru, kellelt milline sõnum tuli.

**Ülevaade ja rakendamine.** Mikservõrgud segavad omavahel mitmest allikast pärinevad sõnumid, muutes sellega konkreetse sõnumi algallika tuvastamise keeruliseks. Lihtsaim sellne mikservõrk koosneb mitmest ahelasse ühendatud vahendusserverist (mikserist), kus iga server eemaldab saabuvatelt sõnumitelt järjestuse ning muudab nende väljanägemist (näiteks krüpteerides sõnumid uue võtmega). Sõnumid suunatakse edasi kas järgmisesse mikservõrgu serverisse või siis nende lõplikku sihtkohta. Serverite omavahelised ühendused määravad ära kasutusel oleva miksimisstrateegia ning mikservõrgu topoloogia. Mikservõrgus on seega vähemalt kaks sõnumite allikat, vähemalt üks miksimist teostav vahendusserver ning vähemalt üks sõnumite sihtkoht.

Mikservõrgud (ingl. k. *mix networks*, *mixnets*) on andmevahetussüsteemid, mis kasutavad erinevaid krüptograafilisi ja permutatsoonivahendeid, et kaitsta delikaatseid andmeid, näiteks tervise- või muid isikuandmeid [129].

Oluline osa mikservõrkude arengust on seotud elektroonilise hääletamisega [130], kus ühelt poolt on oluline tagada anonüümsus, salastatus ning terviklus, kuid samas on eesmärk saavutada ka hääleõiguse korrektne tuvastamine, hääletussüsteemi kasutatavus ning hääletustulemuse korrektsus.

Elektroonilise hääletamise süsteemidest on üks tuntumatest Eesti süsteem. Seda on aastate jooksul pidevalt arendatud ja täiendatud ning valijad on seda ka üha rohkem kasutama hakanud.[52] Välja on toodud ka probleeme, potentsiaalseid turvaauke seoses võltsimise ja muude rünnetega [131]. Pidevalt käigusoleva turvatehnika protsessiga on aga riske järjepidevalt kahandatud.

Teine mikservõrkude kasutusvaldkond on seotud internetipõhiste suhtluskeskkondadega, kus eesmärgiks on tagada kasutajate anonüümsus ja privaatsus, isegi kui suhtlus toimub üle seiratud võrgu. Siit on esile kerkinud anonüümsed e-posti teenused, interneti marsruutimisprotokollid ja digitaalsed vääringusüsteemid, mille kõigi eesmärk vältida tsensuuri. Tavaliselt on tegemist hajussüsteemidega, kus klient-server ühenduse asemel suhtlevad kõik osapooled kui võrdne võrdsega (ingl k *peer-to-peer*).

**Ajaloost.** Esimese anonüümsete sõnumite saatmise süsteemi kirjeldus, mis kasutab mitut järjestikust osapoolt, kes sõnumeid juhuslikult ümber järjestavad ja seeläbi sõnumi allikat varjavad, ulatub tagasi 1980. aastatesse [132].

Seoses suureneva sotsiaalse huviga privaatsuse ja anonüümsuse vastu kogus ka selline välja pakutud süsteem populaarsust ning seda hakati edasi uurima ning arendama. Nüüdseks on mikservõrkudest kujunenud erinevates valdkondades ja riikides kasutatav privaatsuskaitsetehnoloogia, kus nii kasutaja identiteedi kui ka ühenduse metaandmete jälitamatus on võrdlemisi range nõue. Viimaste aastakümnete teadustegevus on lisaks mikservõrkude kontseptsiooni edasiarendamisele keskendunud lisaks ka selle tehnoloogia turvalisuse ja jõudluse tõstmisele ning mikservõrkude topoloogiale.

**Turvagarantiid ja jääkriskid.** Anonüümsus on tagatud sellisel määral, et suhtlust jälgival osapoolel on raske kindlalt kokku viia sõnumi algallikat ja sihtkohta. Oma eesmärgi täitmiseks peavad mikservõrgud olema kaitstud korreleerimisrünnete ja võrguliikluse analüüsi vastu, samuti pahatahtlike või ekslike kasutajate ning vahendusserverite vastu ja lisaks vastu pidama suurele võrgukoormusele. See tähendab, et mikservõrgud on ühtlasi paindlikud ning mastaabitavad. Ühest kompromiteerimata vahendusserverist (st sellisest, mis ei avalda enda kasutatud permutatsiooni) mikservõrgus piisab, et pakkuda sõnumiallika anonüümsust.

Mõnes rakenduses (nt e-hääletamisel) on vaja lisaks tagada, et mikservõrgu vahendusserverid on ausad, st nad ei muuda ega kaota sõnumeid ja ei tekita neid ka juurde. Teisisõnu, iga vahendusserver peab suutma tõestada, et sisse tulnud ja välja antavad sõnumid on täpselt samad, lihtsalt teistsuguse väljanägemisega ja teises järjekorras. Sellised korrektsustõestused realiseeritakse nullteadmustõestuste abil.

[52] Elektroonilise hääletamise statistika https://www.valimised.ee/valimiste-arhiiv/elektroonilise-haaletamise-statistika (viimati külastatud 02.03.2023).

# Mikservõrgud

**SIDE**

**Inglise keeles:** mixnets

**Lühidalt:** Mikservõrk võtab mitu privaatset sõnumit ja segab nad omavahel ära nii, et vaatleja ei saa aru, kellelt milline sõnum tuli.

**Arenduse keerukus:** keskmine

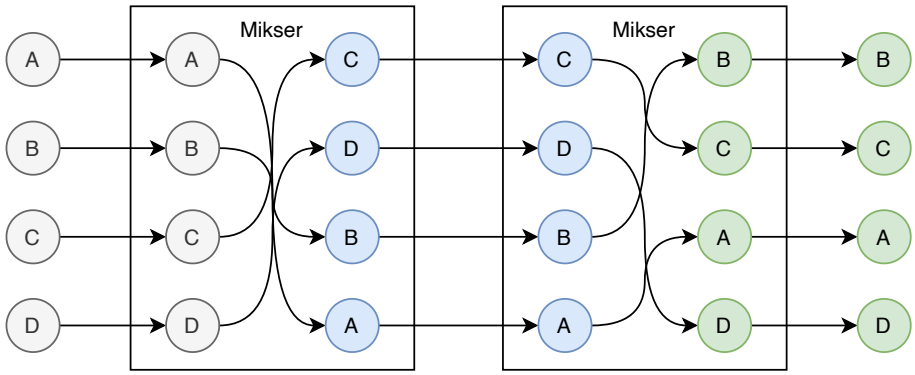
**Ülalpidamise keerukus:** keskmine

**Täpsus:** ei rakendu

**Privaatsusgarantii:** hajutusel põhinev

**Tehnoloogia küpsus:** keskmine

**Ülevaatlik mudel:**




Mikser: sõnumite väljanägemise muutmine (joonisel värvi vahetus) ja juhuslik ümberjärjestamine
⬆ Privaatsus

Mikser
⬆ Privaatsus

**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: võrgus on vähemalt kaks sõnumite algallikat ja vähemalt üks sihtkoht.
2. Turvaeeldus: võrgus on vähemalt üks vahendusserver, mis pole pahatahtlik.
3. Turvaeeldus: tarkvaralised komponendid saavad võrgus kasutada mitmeid erinevaid marsruute.
4. Jääkrisk: võrk on liiga väike või on suur osa sellest vastase kontrolli all.

**Rakendusvõimalused:**

1. Sideseansi allika anonüümsus.

**Õiguspraktika:**

1. Mikservõrgud üksi ei taga side konfidentsiaalsust, seega võivad vahendusserverid sattuda õiguslikkus mõttes töötlema kõikvõimalikke isikustatavaid andmeid.

**Tuntumad rakendused:**

1. E-posti anonüümija Mixminion
2. Eesti internetihääletamise süsteem (jt interneti-hääletuse süsteemid)
3. Anonüümse internetikasutuse platvorm Tor

# Mikservõrgud Eesti internetihääletuses

**Lühidalt:** Eesti internetihääletamise süsteemis kasutatakse mikservõrke, otspunktkrüpteerimist, digiallkirju ja muid krüptograafilisi protokolle, et tagada häälte privaatsus ja mõjutuskindlus.

**Teostamise aasta:** Alates 2005

**Riik:** Eesti

**Omanik:** Riigi valimisteenistus

**Teostaja:** Kuni 2014 oli tarkvara arendaja Cybernetica. Alates 2014 aastast tegeleb sellega Smartmatic Cybernetica Centre of Excellence for Internet Voting (SCCEIV)

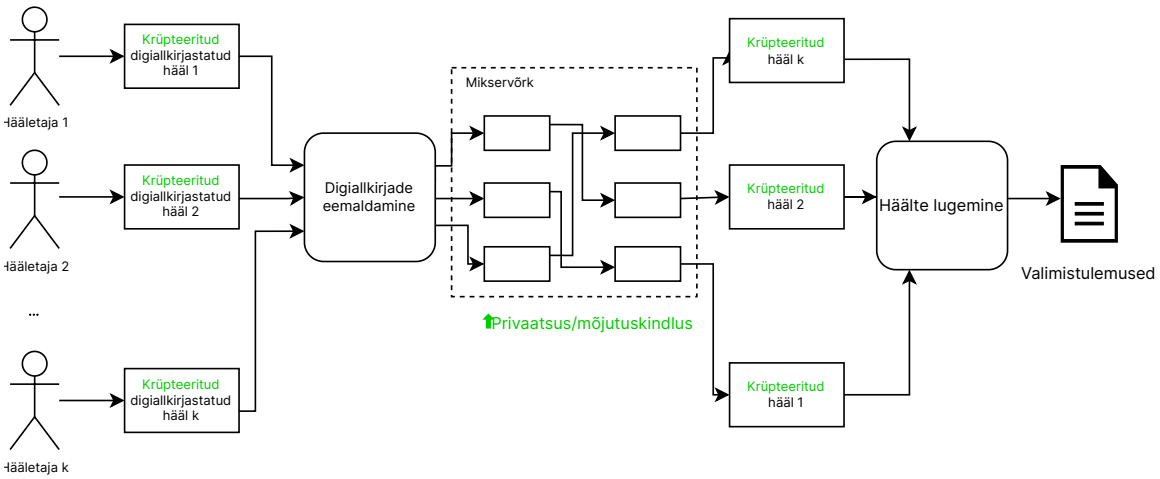
**Süsteemi küpsus:** püsiv juurutus

**Privaatsuskaitse tehnoloogiad:**

1. mikservõrgud
2. otspunktkrüpteerimine

**Sobivad kasutusjuhtumid:**
internetihääletus

**Ülevaatlik mudel:**




**Märkimisväärsed omadused:**

1. Tehnoloogilisest vaatepunktist on Eesti internetihääletamise süsteem üks kaugelearenenumaid usaldatud internetihääletuse keskkondi.
2. Alates 2005. aastast on internetihääletamist kasutatud mitmetel kohaliku omavalitsuse, Riigikogu ja Euroopa parlamendi valimistel. Internetihäälte osakaal valimistel on ajas pidevalt kasvanud, 2019 aasta Riigikogu valimistel moodustasid internetihääled kõigist antud häältest 43.8%, 2021 aasta kohaliku omavalitsuse volikogu valimistel 46.9%.
3. Eksperdid Eestist kui ka välismaalt on süsteemi auditeerinud ja uurinud, eriti keskendudes turvalisuse, privaatsuse, tervikluse ja mõjutuskindluse aspektidele.

### 4.4.3 Sibulmarsruutimine

**Lühidalt.** Sibulmarsruutimine (ingl k *onion routing*) aitab peita saadetava sõnumi saatjat ja saajat võrgu jälgija eest.

**Ajaloost.**

Sibulmarsruutimine on protokoll, mis võimaldab anonüümset suhtlust üle (avaliku) arvutivõrgu [133].

Oma ülesehituselt sarnaneb see mikservõrkudele, kuna tekitab avaliku võrgu peale oma virtuaalse kihi eesmärgiga anonüümida selles toimuvat sõnumivahetust. Sibulmarsruutimise esmane kasutus on anonüümne veebi sirvimine ja marsruutimine üldisemalt, et kaitsta inimesi seiramise eest või pääseda mööda tsensuurist. Samas, lisaks kasutatakse tehnoloogiat ka masinate kaughalduseks, failide jagamise rakendustes ning ka suhtlusvõrgustikes. Üldisemalt, kuna sibulmarsruutimine pakub anonüümsust võrguühenduse tasemel, siis saab selle peale ehitada mitmeid tänapäevaseid rakendusi.

**Ajaloost.** Tehnoloogia pakkusid esmakordselt 1990. aastate keskel välja USA luureagentuurid, et kaitsta oma suhtlust üle interneti. Hiljem kasutasid seda ka teised riigiasutused ning 2002. aastal avalikustati see tehnoloogia avatud lähtekoodiga TOR (*The Onion Router*) projekti osana [134].

**Ülevaade ja rakendamine.** Sibulmarsruutimise protokoll koosneb kahest osast: andmestruktuurist ja marsruutimismehhanismist. Andmestruktuuriks on võrgupakett, mis on mässitud mitme erinevate võtmetega krüpteeringukihi sisse (sellest ka tehnoloogia nimi).

Marsruutimismehhanism koosneb mitmest vahendusserverist, marsruuterist, mis valitakse suurest vahendusserverite hulgast, enamasti juhuslikkuse alusel. Valitud vahendusserverid moodustavad ahela, kus iga lüli krüpteeritud võrgupakette vastu võtab, töötleb ja järgmisesse marsruuteerisse edasi suunab. Need vahendusserverid teavad vaid vahetult eelmist ja järgmist sammu võrgupaketi teekonnas ning iga sellise vahendusserveri ülesanne on võrgupaketilt eemaldada üks kiht krüpteeringut ja saata see ahelas järgmise töötlejani kuniks see lõpuks oma sihtkohta jõuab.

Sibulmarsruutimine pakub reaalajas kahesuunalist võrguliikluse anonüümimist, peites lõppkasutajate võrguühenduse metaandmed nagu IP-aadressid ja geograafilise asukoha. Sarnaselt on võimalik peita ka teenuseandja serveri IP aadressi (ja seega geograafilist asukohta), mis paneb aluse pimevõrgu toimimisele.

**Turvagarantiid ja jääkriskid.** Sibumarsruutimise tehnoloogia võimaldab anonüümida võrguliiklust ja seeläbi piirata kasutajate isikustamist. Enamik selle tehnoloogia pakutavatest garantiidest ja puudujääkidest on otseselt seotud saadaval olevate vahendusserverite arvuga ning valitud vahendusserverite ahela pikkusega. Kuna sibulmarsruutimine on üles ehitatud hajusa detsentraliseeritud võrguna, pakub see lisaks tõhusat kaitset tsensuuri vastu, mastaabitavust ning tõrkekindlust.

Samas, olenevalt valitud ahela pikkusest ja sinna sattuvatest vahendusserveritest on võrgu läbilaskevõime piiratud. Potentsiaalsete desanonüümimisrünnete või pahatahtlike vahendusserverite tõttu ei ole sibulmarsruutimise poolt pakutav kaitse alati garanteeritud, aga mida rohkem on sõltumatuid vahendusservereid, seda väiksem on selliste rünnete õnnestumise tõenäosus.

Tor[53] on vabavaraline ja avatud lähtekoodiga tarkvara, mille eesmärk on anonüümida kasutaja suhtlust üle interneti ja piirata selle jälitatavust sibulmarsruutimise kasutamisel. See lubab varjata tarkvara kasutaja (sõnumi saatja) kui ka Tori võrgus oleva teenuseandja võrguliikluse metaandmeid, identiteeti ja asukohta.

Suure osa kasutajatest moodustavad aktivistid, ajakirjanikud ja muud isikud riikidest, kus interneti kasutust piiratakse [135].Samas leiab see laialdast kasutust ka kurjategijate seas, mille tõttu on teenust kritiseeritud ja mõnes riigis täielikult keelatud [136].

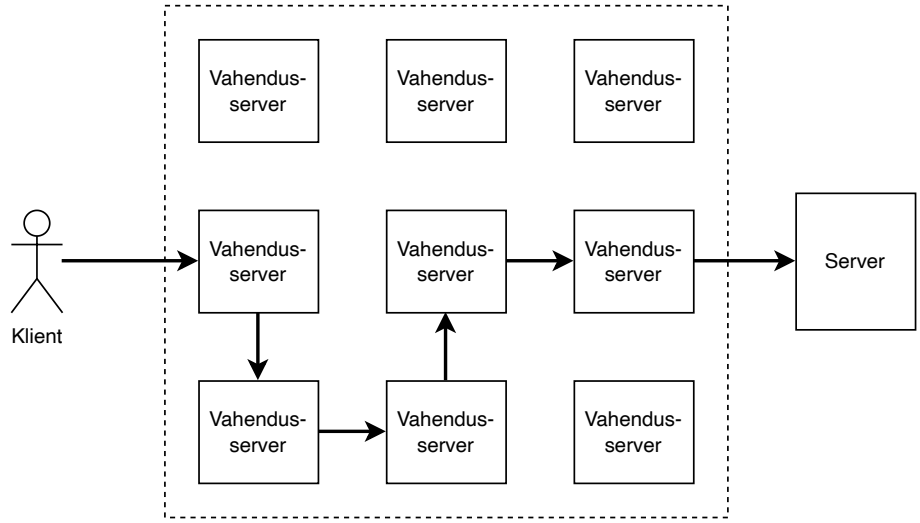
**Juhised rakendajale.** Sibulmarsruutimise kasutamine rakendustes võib olla väga lihtne, kui selleks kasutada olemasolevaid võrke nagu Tor. Leidub ka teeke, mille abil saab rajada omaenda sibulmarsruutimise võrke. Kuid see on keerukas ning kuskilt tuleb leida selle võrgu jaoks piisav arv vahendusservereid.

**Õiguslikud aspektid.** Mõnes riigis väljaspoolt Euroopat on näiteks Tor võrgu liiklus blokeeritud [137]. Mõnes riigis võib Tor vahendusserveri (sõlme) majutaja sattuda korrakaitse huviorbiiti[54].

[53] https://www.torproject.org (viimati külastatud 02.03.2023).
[54] The Legal FAQ For Tor Relay Operators https://community.torproject.org/relay/

| **Sibulmarsruutimine** | **SIDE** |
|---|---|
| **Inglise keeles:** onion routing | |
| **Lühidalt:** Sibulmarsruutimine aitab peita saadetava sõnumi saatjat ja saajat võrgu jälgija eest. | **Arenduse keerukus:** kõrge |
| | **Ülalpidamise keerukus:** kõrge |
| | **Täpsus:** — |
| | **Privaatsusgarantii:** krüptograafial ja hajutusel põhinev |
| | **Tehnoloogia küpsus:** kõrge |

**Ülevaatlik mudel:**




Sibulmarsruutitud võrgus näeb iga osapool enda naabreid, aga keegi ei näe paketi sisu
⬆ Privaatsus
⬇ Kiirus

**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: vahendusserverite võrk on piisavalt suur, et sealt konkreetseks sideseansiks valitud ahel oleks piisavalt juhuslik.
2. Turvaeeldus: tuleb välistada, et mõnel vastasel on kogu võrgu üle kontroll nii, et ta suudab võrgu sisenevat ja väljuvat liiklust korrelleerida.
3. Jääkrisk: võrk on liiga väike on liiga suur osa sellest ründaja kontrolli all.

**Rakendusvõimalused:**

1. Sideseansi allika ja sihtkoha anonüümsus avalikus sidevõrgus

**Õiguspraktika:**

1. Sõltuvalt võrgu seadistusest ja kasutatavast teenusest võivad vahendusserverid sattuda õiguslikkus mõttes töötlema kõikvõimalikke isikustatavaid või muud sorti andmeid, millega nad tingimata seotud ei ole.

**Tuntumad rakendused:**

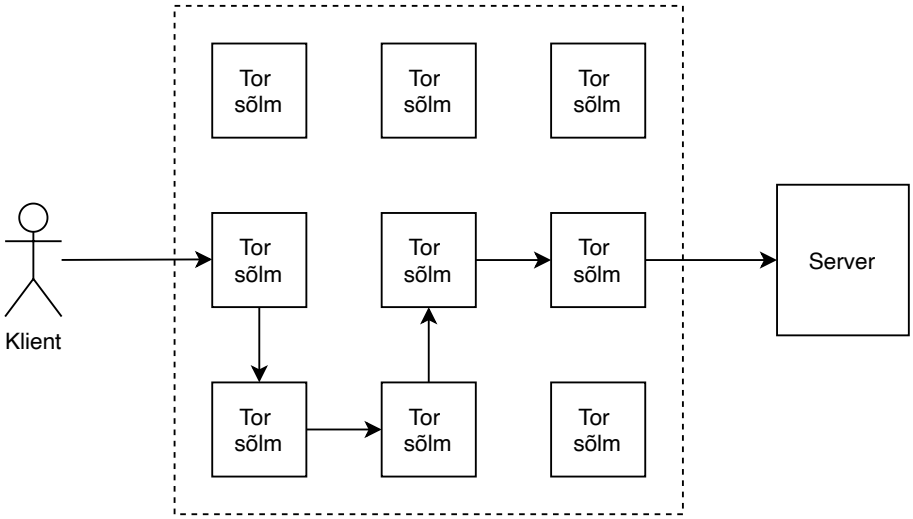
1. Anonüümne veebi sirvimine (nt Tor)
2. Peidetud veebiteenused ja suhtluskeskkonnad, mille geograafiline asukoht pole teada (nn pimeveeb)
3. Anonüümne failijagamine (nt Tribler)

`community-resources/eff-tor-legal-faq` (viimati külastatud 01.03.2023).

## Sibulmarsruutimise platvorm Tor

| | |
|---|---|
| **Lühidalt:** Tor on ülemaailmse kasutajaskonnaga sibulmarsruutimise platvorm. | **Teostamise aasta:** Alates 2002 |
| | **Riik:** Globaalne ja detsentraliseeritud |
| | **Omanik:** omanik puudub, opereerijaks on vabatahtlikud sibulmarsruuterid |
| | **Teostaja:** välja töötatud USA mereväe teaduslaboris, kuid nüüd hooldab ja arendab peamiselt mittetulunduslik organisatsioon The Tor Project, Inc, mida veavad arvutiteadlased |
| | **Süsteemi küpsus:** püsiv juurutus |
| **Privaatsuskaitse tehnoloogiad:**<br>1. sibulmarsruutimine | **Sobivad kasutusjuhtumid:**<br>anonüümne internetikasutus |

**Ülevaatlik mudel:**




Tor võrk: iga sõlm näeb enda naabreid, keegi ei näe paketi sisu
⬆ Privaatsus
⬇ Kiirus

**Märkimisväärsed omadused:**

1. Internetis kasutajate ja teenusepakkujate privaatsuse kaitsmiseks ja tsensuuri vältimiseks mõeldud tarkvaralahenduste seas on Tor üks laialdasemalt kasutatavaid.
2. Tor on võitnud mitmeid auhindu. Näiteks Free Software Foundation's 2010 Award for Projects of Social Benefit, millega tunnustati seda kui ühiskondlikke hüvesid pakkuv vabavaralist tarkvara, ja the 2012 EFF Pioneer Award.
3. Toril on hinnanguliselt kaks miljonit igapäevast kasutajat (Eestis 1500).

# 4.5 Läbipaistvust toetavad tehnoloogiad

## 4.5.1 Läbipaistvuse tehnoloogiate iseärasused

Läbipaistvus privaatsuskaitse eesmärgina taotleb kogu privaatsuse seisukohalt relevantse andmetöötluse arusaadavuse ja rekonstrueeritavuse garanteerimist igal ajahetkel kõigile kaasatud osapooltele (näiteks andmetöötleja, inimene andmete omanikuna, audiitor, järelvalveasutus jne) [138]. Läbipaistvuse nõue isikuandmete töötlemisel tuleneb ka otseselt IKÜM-ist.

Asjakohane informatsioon andmetöötluse kohta peab olema saadaval enne andmetöötluse algust, selle kestel ja pärast seda. Kuna läbipaistvuse eesmärk on suurendada arusaamist andmetöötlusest ja sellega seotud riskidest, sõltub info sihtauditooriumi võimekusest, mis ulatuses seda pakkuda ja kuidas seda edasi anda. Seepärast on läbipaistvuse mehhanismide teabe levitamise kanalite kujundamisel oluline pöörata tähelepanu kasutatavusele, hõlpsusele ja kaasatusele [139]. Mitte kõiki inimesi ei huvita detailne sissevaade andmetöötlusesse, teisi aga ei rahuldaks vaid üldistatud informatsioon. Seetõttu soovitatakse siin informatsioon kättesaadavaks teha mitmel detailsustasandil.

Läbipaistvus võiks ideaalis olla juurutatud süsteemi osana juba selle kavandamisel, ent on võimalik ka tagantjärele, integreerides lahendused, mis tagavad soovitud läbipaistvustaseme saavutamiseks kogu vajaliku informatsiooni talletamise, seostamise, agregeerimise ja väljastamise. Oluline on, et sellised teenused oleksid sihtauditooriumile kasutataval ja arusaadaval kujul kättesaadavad. Näiteks ei tohiks eeldada, et inimene soovib konkreetse privaatsusintsidendi tuvastamiseks ise läbi uurida suuremahulist andmekogu, vaid tema toetamiseks on tarvis luua otsingufunktsioon.

## 4.5.2 Dokumenteerimine, logimine ja osapoolte teavitamine

Läbipaistvuse saavutamine võib toimuda andmetöötluse dokumenteerimise, logimise ja raporteerimise, kasutajateavituste jmt abil [1]. Dokumenteerimise võimaluste spekter on lai, ulatudes organisatsiooniplaanidest üle süsteemiarhitektuuri käsiraamatute kuni lähtekoodi liideste kirjeldusteni, mis kõik panustavad läbipaistvuse tagamisse. On oluline, et kõik andmetöötlussüsteemi aspektid oleksid määratletud etteulatuvalt, nii et selles ei saa kunagi tekkida dokumenteerimata seisundit või realiseeruda mõni ettenägematu toiming.

Dokumentatsiooni kõrval tagavad läbipaistvuse spetsiifilised ja terviklikud **logimismehhanismid**, mille ülesanne on näiteks kliendiinteraktsiooni toimudes selle kohta logisüsteemi kirjete kandmine. Tuleb tähele panna, et tagades ühte privaatsuseesmärki - läbipaistvust - on logidesse koguneva informatsiooni suhtes teise privaatsuseesmärgi - seostamatuse - tagamiseks tarvis rakendada teisi privaatsuskaitse tehnoloogiaid.

Olulist rolli omavad ka mehhanismid talletatud informatsiooni toimetamiseks asjakohastele osapooltele (teavitamiseks). Teavitada tuleb inimest, kelle andmeid töödeldakse, aga ka organisatsiooni juhtkonda, järelvalve teostajat, partnereid vastavalt sõlmitud lepingutele ja kohustustele jne. Selline informeerimine saab toimuda päringust tulenevalt või proaktiivsete teavitustena toimunud andmetöötlussündmuste kohta. Läbipaistvust parandab ka **kasutajatoe tugiteenus**, kui see oskab vastata andmetöötlust puudutavatele küsimustele.

Privaatsust puudutavate aspektide hulka, mille osas läbipaistvust tuleks taotleda [140], kuuluvad:

1. kogutud andmed, sh kelle kätte ja missuguste tingimuste alusel;
2. andmevoog – nii plaanitud kui tegelik ja sellega seotud riskid;
3. privaatsuspoliitika;
4. andmetöötluse meetodid;
5. (andmetöötluse tulemusel) pakutavad teenused;
6. kasutatud tarkvara ja tehnoloogiad;
7. interaktsioonipartnerite reputatsioon;
8. kogu andmetöötluse usaldusväärsuse ja turvalisuse garantiid;
9. kõik tõelised või võimalikud nõrkused ja turvaintsidendid.

### 4.5.3 Kasutustingimuste arusaadavuse parandamine

Interneti andmesidevõrgu plahvatusliku leviku kontekstis 1990-datel hakkasid üha enamad organisatsioonid oma kodulehtedel avaldama kasutustingimusi. Nende dokumentide keerukus peegeldas informatsioonipraktikate üha suurenevat keerukust. Kasutajate jaoks olid need dokumendid segadust tekitavad ning sai selgeks, et need ei anna inimesele selget ülevaadet, kuidas tema andmeid kasutatakse. Organisatsioonid olid hädas vastakate eesmärkidega luua kasutustingimusi, mis oleksid lihtsad mõista, terviklikud ja vastaksid ka regulatsioonidele. 2003. aastal 25. Rahvusvahelisel andmekaitsekonverentsil Sydney's tõsteti lahendusena esile **mitmekihiliste kasutustingimuste** loomist. Järgmisel aastal koostasid peamiselt Euroopa andmekaitseinspektorid, teiste valitsusametite esindajad ning Euroopa tarbijakaitsening ärijuhid mitmekihiliste kasutustingimuste põhimõtteid kirjeldava Berliini memorandumi. Sama aasta lõpus andis oma heakskiidu ka direktiivi 95/46/EÜ artikli 29 alusel asutatud andmekaitse töörühm (*Article 29 Working Party*, edaspidi ärtikli 29 töörühm").

Teenuste kasutustingimused on traditsioonilist tähendanud kümneid lehekülgi juriidilises žargoonis teksti, mille sisuga tavakasutaja üldjuhul ei tutvu. Selles väljendub sisemine võimuasümmeetria andmetöötleja ja inimese kui andmesubjekti vahel: kasutustingimused on koostatud eelkõige andmetöötleja huvide kaitseks ja riskide leevendamiseks, need ei ole paindlikud ega üldjuhul keskmise kasutaja jaoks arusaadavad [141]. Õiguslikult siduva dokumendina on juriidiline keerukus osaliselt vältimatu, ent sisu mõistetavuse suurendamiseks on võimalik selle olulisemaid osasid esile tõsta. Seda võib teha nii teenusepakkuja ise või sõltumatu osapool. Seoses GDPR-i jõustumisega 2018. aastal, mis nõuab kasutustingimustelt kompaktsust, läbipaistvust ja arusaadavust, on selles liinis toimunud parenemine, sh on paljud organisatsioonid juurutanud allpool kirjeldatud lahendusi.

Läbipaistvus eeldab kasutustingimusi, mis on hõlpsad mõista, soodustavad võrdlust ning mille pinnalt on osapooltel võimalik tegutseda. Selleks, et kasutustingimused oleksid kergesti loetavad ja mõistetavad, peavad need olema lühikesed, ning olema esitatud lihtsas keeles ja vormis. Terviklikud kasutajatingimused kipuvad olema pikad ja keerukamad, nii et pole võimalik ühe dokumendiga mõlemaid kriteeriume täita. Mitmekihilised kasutajatingimused koosnevad kokkuvõtlikest tingimustest, mis sisaldavad lihtsalt mõistetaval ja kasutataval kujul kõiki võtmekomponente, ning terviklikke kasutajatingimusi kõigi vajalike juriidiliste osadega. Need suurendavad kasutajate usaldust andmetöötleja suhtes. Uuringud näitavad, et kasutajad eelistavad kokkuvõtlikke kasutustingimusi täistekstidele, kuna neile sobib informatsioon, mis on esitatud selgelt, visuaalselt atraktiivselt ning hõlpsasti võrreldavalt. Samuti võimaldab see rahvusvahelistel organisatsioonidel esitada oma kasutustingimusi riigist riiki ühelaadselt.

Artikli 29 töörühm soovitab maksimaalselt kolmekihilist [142] struktuuri:

- Esimene kiht – lühitingimused: minimaalne komplekt väga piiratud suurusega alal (nt mobiilirakendustes) kasutamiseks, mis annab infot näiteks vaid andmetöötleja ja tema kontaktide ning andmetöötluse eesmärkide kohta;
- Teine kiht – kokkuvõtlikud tingimused: põhitingimuste ülevaade vähem kui ühel leheküljel, mis ideaalis kasutab alapealkirjasid ja katab tingimuste geograafilist ulatust, kogutavaid isikuandmeid, andmete kasutamist ja jagamist, valikuid (nt andmetele ligipääsetavus), olulist lisainfot ja kontaktandmeid;
- Kolmas kiht – terviklikud kasutustingimused oma õiguslikus täielikkuses.

Heaks näiteks mitmekihilistest kasutajatingimustest on LinkedIni kasutajaleping, milles olulisemad põhimõtted on tekstist kõnekeelsete kokkuvõtetena välja toodud[55]. Teenusepakkujatest sõltumatult on kasutajatingimusi hakanud analüüsima ja lihtsustama kogukonnad nagu ToS;DR[56] ja TOSBack[57].

Kirjeldatud esimese kihi alla võib liigitada ka hiljuti nii iOS kui Android rakenduste ökosüsteemides juurutatud **privaatsusmärgised** (*privacy nutrition labels*)[58]. Nende eesmärk on anda standardiseeritud kujul kokkuvõtlik ülevaade andmepraktikatest, mis inimestele tavaliselt muret teevad. Uuringud on siiski tuvastanud, et oma esialgsel kujul kaasnevad privaatsusmärgistega vääritimõistmised ja rahulolematus [143].

[55] LinkedIn User Agreement https://www.linkedin.com/legal/user-agreement (viimati külastatud 23.02.2023).
[56] ToS;DR https://tosdr.org/en/about (viimati külastatud 23.02.2023).
[57] TOSBack https://tosback.org (viimati külastatud 23.02.2023).
[58] Apple Privacy Nutrition Labels https://www.apple.com/privacy/labels (viimati külastatud 23.02.2023).

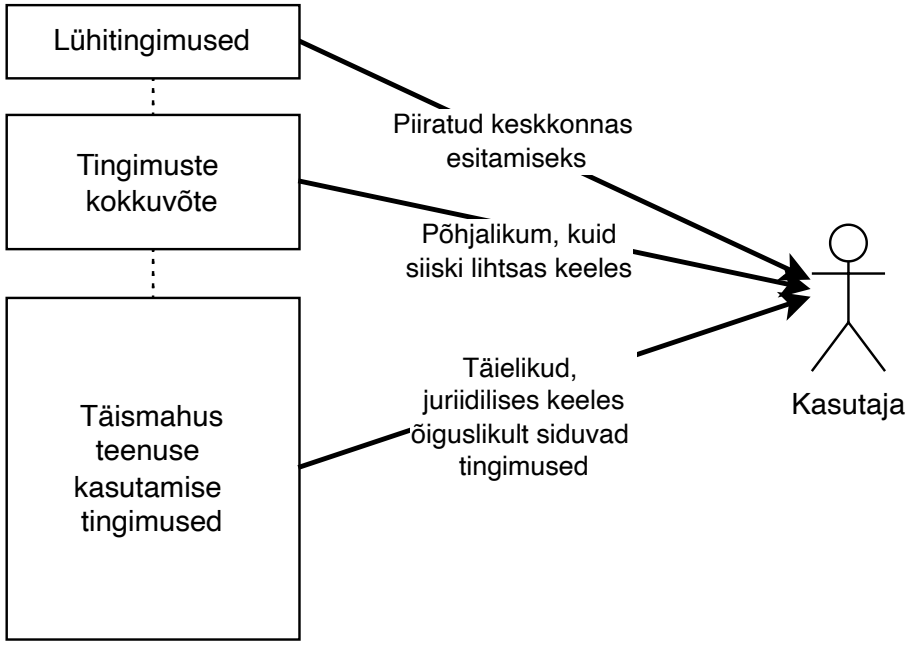
Kokkuvõtlikult võib tõdeda, et ehkki universaalselt mõistetavast märgistusest (umbes nagu liiklusmärgid) oleme veel kaugel, on privaatsusmärgised siiski lootustandev lahendus andmetöötluse läbipaistvuse suurendamisel. Privaatsusmärgiste parendamiseks vajalikest põhimõtetest annab ülevaate Cranor [144].

**Täppisteatised** ilmuvad ekraanile andmete sisestamise ajal ning informeerivad inimest *ad hoc*, kuidas sisestatavaid andmeid kasutatakse – just sellal, kui inimese jaoks on kõige relevantsem aeg selle infoga tutvumiseks. Need on eriti tõhusad olukorras, kus inimene jagab isikuandmeid näiteks ostu või muu interaktsiooni erinevates punktides – sageli organisatsiooni kodulehel vorme täites. Selles olukorras inimene tavaliselt ei mõtle, mis nendest andmetest toimingu sooritamise järgselt saab, samas kui täppisteatised tagavad selle info sobivates ühikutes inimeseni toomise, mis on tunduvalt tõhusam kui pikkadele kasutajalepingutele toetumine, mille kontekstiväline mõistmine on keerukas. Eriti tõhusad on täppisteatised kombinatsioonis teiste tehnikatega, tagamaks et rohkem informatsiooni soovivatel inimestel on selle hankimine antud olukorras lihtne. Näiteks võib täppisteatisega kaasuda link enama informatsiooni juurde.

**Andme- ja nõusolekukviitungid** teenivad tarbijate vajadust lihtsate privaatsusväljavõtete järele, mis selgitavad, kuidas nende andmeid ja nõusolekuid kasutatakse, toetudes traditsioonilisele ostutšeki analoogiale. Uuringutes on demonstreeritud, et andmekviitungi juurutamine lisab isikuandmete kasutamisse lisa-läbipaistvuskihi, mis toetab usalduse kujunemist organisatsiooni ja selle klientide vahel [145]. Eriti just nõusoleku mõistmise puhul on traditsiooniliselt tegu olnud väga kompleksse interdistsiplinaarse probleemiga - iga veebipõhine nõusoleku juurutamine peab vastama kasutatavuse-, juriidilistele, tehnilistele ja ärinõuetele. Selle keerukuse tõttu on senised pingutused peamiselt keskendunud elementaarse regulatsioonidega vastavuses olemisele ja nõusoleku automatiseerimisele ning vähem inimestele oma andmete üle tõelise kontrolli tagamisele. Nõusolekukviitung pakub nõusolekusse puutuvate interaktsioonide talletamiseks uue ent tuttavatel alustel paradigma. Andmetöötleja seisukohalt tagab nõusolekukviitung nõusoleku kehtivuse demonstreerimise, samal ajal kui inimese jaoks pakub see läbipaistvust ja tõendusmaterjali võimalike nõusoleku väärkasutamise juhtude olukorras ning nõusolekust tulenevate õiguste realiseerimisel [146]. Andme- ja nõusolekukviitungeid võib kasutada eraldi, paralleelselt või omavahel seostatuna.

| **Mitmekihilised kasutustingimused** | **LÄBIPAISTVUS** |
|---|---|
| **Inglise keeles:** Layered notice form | |
| **Lühidalt:** Mitmekihilised kasutustingimused on arusaadavad nii tavalisele inimesele kui ka neid koostanud juristile. | **Arenduse keerukus:** madal<br>**Ülalpidamise keerukus:** madal<br>**Täpsus:** —<br>**Privaatsusgarantii:** organisatoorne lubadus<br>**Tehnoloogia küpsus:** kõrge |

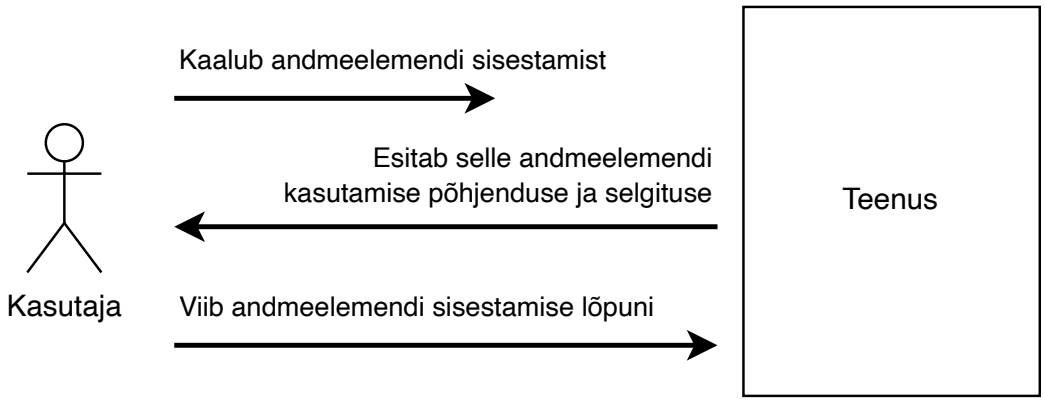
**Ülevaatlik mudel:**




| | |
|---|---|
| **Turvaeeldused ja jääkriskid:**<br>1. Turvaeeldus: kasutustingimused ja poliitikad on jõustatud organisatsiooniliste ja tehniliste meetmetega.<br>2. Jääkrisk: andmeid töödeldakse poliitikate vastaselt. | **Rakendusvõimalused:**<br>1. Kõik teenused, mis avaldavad kasutustingimusi. |
| **Õiguspraktika:**<br>1. Artikli 29 töörühm on soovitanud lihtsas keeles mitmekihilisi kasutustingimusi IKÜM teavitusnõuete täitmiseks. | **Tuntumad rakendused:**<br>• LinkedIni kasutajaleping<br>• Lihtsustatud tingimuste teenused ToS;DR ja TOS-Back<br>• Apple Privacy Nutrition Labels |

| **Täppisteatised** | **LÄBIPAISTVUS** |
|---|---|
| **Inglise keeles:** Just-in-time notices | |
| **Lühidalt:** Täppisteatis ilmub ekraanile täpselt andmete sisestamise ajal ning ütleb, mida nende andmetega tegema hakatakse. | **Arenduse keerukus:** madal<br>**Ülalpidamise keerukus:** madal<br>**Täpsus:** —<br>**Privaatsusgarantii:** organisatoorne lubadus<br>**Tehnoloogia küpsus:** kõrge |

**Ülevaatlik mudel:**




| **Turvaeeldused ja jääkriskid:** | **Rakendusvõimalused:** |
|---|---|
| 1. Turvaeeldus: kasutustingimused ja poliitikad on jõustatud organisatsiooniliste ja tehniliste meetmetega.<br>2. Jääkrisk: andmeid töödeldakse poliitikate vastaselt. | 1. Kõik teenused, mis koguvad elementhaaval andmeid. |
| **Õiguspraktika:**<br>1. Suurbritannia andmekaitsja ICO on täppisteatiseid populariseerinud IKÜM nõuete täitmise abivahendina. | **Tuntumad rakendused:**<br>— |

## Nõusoleku andmise kviitungid

**LÄBIPAISTVUS**

**Inglise keeles:** Consent receipts

**Lühidalt:** Nõusoleku andmise kviitung meenutab andmesubjektile, kellele ta mille jaoks nõusoleku andnud on.

**Arenduse keerukus:** madal

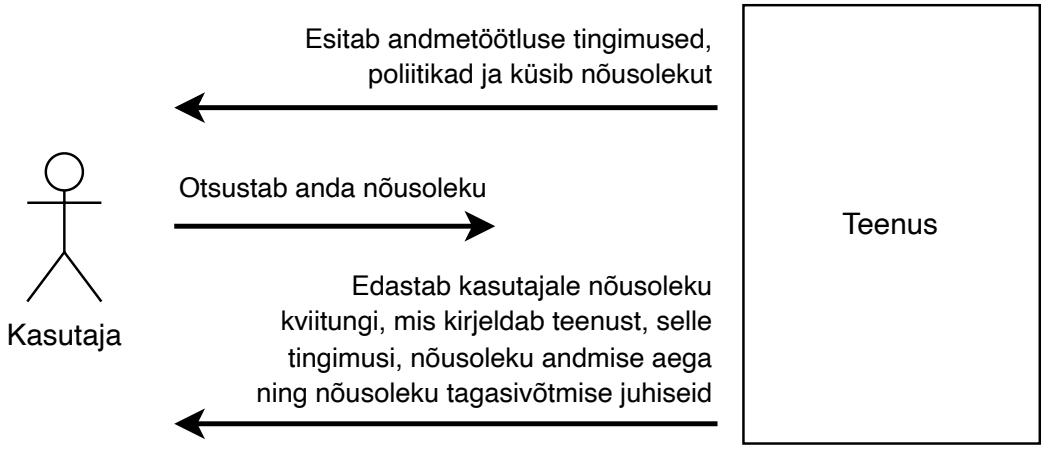
**Ülalpidamise keerukus:** madal

**Täpsus:** —

**Privaatsusgarantii:** organisatoorne lubadus

**Tehnoloogia küpsus:** keskmine

**Ülevaatlik mudel:**




**Turvaeeldused ja jääkriskid:**

1. Turvaeeldus: nõusoleku eelduseks olevaid poliitikaid ja tingimusi jõustatakse organisatsiooniliste ja tehniliste meetmetega.
2. Jääkrisk: andmeid töödeldakse poliitikate vastaselt.

**Rakendusvõimalused:**

1. Kõik teenused, mis koguvad kasutajalt nõusolekut, võivad väljastada selle kohta kviitungi.

**Õiguspraktika:**

—

**Tuntumad rakendused:**

—

# 4.6 Sekkutavust toetavad tehnoloogiad

## 4.6.1 Sekkutavuse tehnoloogiate iseärasused

Sekkutavus tähendab süsteemi omadust võimaldada sekkumist kõikvõimalikku hetkel toimuvasse või planeeritud privaatsuse seisukohalt relevantsesse andmetöötlusesse [138]. Eelkõige tähendab see inimese (andmesubjekti) võimalust andmetöötlusesse vahetult muudatusi teha ning korrektuure sisse viia. IKÜMi kontekstis hõlmab sekkutavus andmesubjekti õiguseid, andmete ülekantavust ning teisi kontrollimeetmeid nagu näiteks nõusolek.

Sealhulgas teenib sekkutavus inimese õiguseid:

1. katkestada teda puudutavate andmete edastamine kolmandale osapoolele;
2. andmete korrigeerimiseks, tagasivõtmiseks ja kustutamiseks;
3. nõusoleku tagasivõtmiseks ja
4. andmetöötluse või selle tulemuse vaidlustamiseks.

Sekkutavus on oluline ka teiste osapoolte jaoks – näiteks olukorras, kus kolmanda osapoole pilvteenust kasutav ettevõte peab kustutama oma (üksikisikust) kliendi isikuandmed.

Sekkutavuse juurutamine loob süsteemi arendusse nii funktsionaalseid kui mittefunktsionaalseid lisanõuded. Süsteem peab olema piisavalt töökindel, tulemaks toime kasutajate sekkumisest tingitud andmete osalise kättesaadamatuse. Süsteemi talitlus peab kohanema andmete eemaldamisega kasutaja soovil. Parajasti toimivatesse (andmetöötlus)protsessidesse sekkumise võimaluse välja ehitamine. Sõltuvalt rakenduskohast võib see olla võimalik nõusoleku tagasivõtmisega.

### 4.6.1.1 Privaatsus- ja andmetöötluspaneelid ning iseteenindused

Privaatsuspaneelid on konkreetse teenuse või rakenduse kasutajaliidese komponendid, mis annavad ülevaate, missuguseid andmeid kogutakse, kuidas neid kasutatakse, missuguste kolmandate osapooltega jagatakse jne. Niisuguste paneelide kujundamisel on vaja pöörata tähelepanu, et kasutajaid mitte eksiteele juhtida, et nad oma seadistusi kohandades soovitud privaatsuseesmärkidest hoopis ei eemalduks. Kuna informatsioon on esitatud deklaratiivsel moel, on peab kasutaja otsustama, kas usaldab, et teenusepakkuja esitab andmetöötlust ja privaatsust puudutavat situatsiooni ausal ja ammendaval moel.

Paljude organisatsioonide puhul moodustavad privaatsus- ja andmetöötluspaneelid osa üldisemast veebipõhisest iseteeninduskeskkonnast, mis loovad võimaluse tutvuda enda kohta kogutud andmetega ning neid teatud ulatuses hallata, näiteks oma andmeid mujal kasutamiseks alla laadida, täites seega IKÜM nõudeid oma andmetest koopia saamiseks.

**Näited privaatsuspaneele ja iseteeninduskeskkondasid hõlmavatest teenustest**

- Eesti Andmejälgija[59]
- Eesti Maksu- ja Tolliameti iseteenindus
- Patsiendiportaal (Eesti Tervise Infosüsteem)
- Google Andmed ja privaatsuspaneel
- Android 12 *Privacy Dashboard* paneel (*Settings* → *Privacy* → *Privacy Dashboard*) kuvab, missugused rakendused on neile antud ligipääsuõigusi viimati kasutanud
- Facebook *General Account Settings* paneel
- Apple *Data and Privacy* paneel

### 4.6.1.2 Dünaamiline nõusolekute haldus

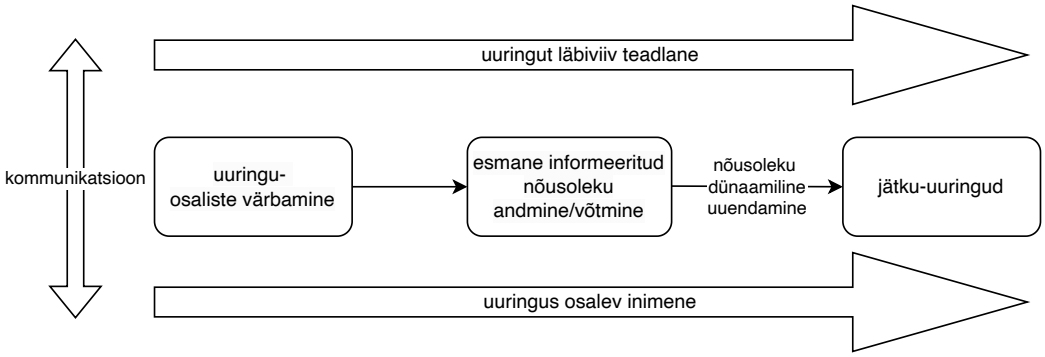
Nõusoleku kasutamine andmetöötluse alusena tähendab inimesele tõelise valiku ja kontrolli tagamist. Tõeline nõusolek peaks andma inimesele otsustusõiguse ja -võimaluse, toetama usalduse ja osalemise

[59] Andmejälgija https://www.ria.ee/riigi-infosusteem/inimkeskne-andmehaldus/andmejalgija (viimati külastatud 23.02.2023).

kasvu inimese ja andmetöötleja vahel ning positiivselt mõjutama andmetöötleja reputatsiooni. Nõusolek peab olema ühetähenduslik ja hõlmama inimese tehtava valiku ühemõttelist kinnitust. Nõusoleku tagasivõtmine peab olema sama hõlbus nagu selle algne andmine. Selleks sobivad eelkirjeldatud privaatsus- ja andmetöötluspaneelid, kus kasutaja saab oma eelistusi muuta. Andmetöötleja IT-süsteemi ülesehitusest ja haldusprotsessidest sõltub, kas andmesubjekti nõusoleku tagasivõtmine katkestab andmetöötluse vahetult või viiteajaga.

| **Nõusolekute haldus (dünaamiline nõusolek)** | **SEKKUTAVUS** |
|---|---|
| **Inglise keeles:** Dynamic consent | |
| **Lühidalt:** Dünaamiline nõusolek lubab kasutajal lihtsalt muuta oma varasemaid nõusolekuotsuseid. | **Arenduse keerukus:** keskmine |
| | **Ülalpidamise keerukus:** madal |
| | **Täpsus:** — |
| | **Privaatsusgarantii:** tõestatavat privaatsusgarantiid ei ole |
| | **Tehnoloogia küpsus:** organisatoorne lubadus |

**Ülevaatlik mudel:**




| **Turvaeeldused ja jääkriskid:** | **Rakendusvõimalused:** |
|---|---|
| 1. Turvaeeldus: andmetöötluse jaoks nõusoleku võtnud osapool peab tagama nõusoleku tingimuste täitmise.<br>2. Turvaeeldus: kui nõusoleku andja esitab uue tahteavalduse, siis peab nõusoleku võtja tagama selle täitmise.<br>3. Jääkrisk: andmeid töödeldakse poliitikate või tahteavalduse vastaselt. | 1. Juurutatav kõikvõimalikes nõusolekupõhistes süsteemides, mitte vaid teaduslikes. |
| **Õiguspraktika:** | **Tuntumad rakendused:** |
| 1. Nõusolek peab olema spetsiifiline, ent teadusuuringutes kasutatakse ka nn ”laia nõusoleku”lähenemist, mille õiguspärasuses ei saa täielikult kindel olla. Dünaamiline nõusolek lahendab selle probleemi, võimaldades nõusoleku andmist erineval üldistustasemel. | 1. Eesti nõusolekuteenus (Riigi Infosüsteemi Amet)<br>2. Eleringi Estfeedi nõusolekuteenus<br>3. TEHIKu nõusolekuteenus |

# 5 Teiste riikide kogemused ja rakendused

## 5.1 Ameerika Ühendriigid

### 5.1.1 Ameerika Ühendriikide ja Ühendkuningriigi PET programm

2022. aastal algas Ameerika Ühenriikide ja Ühendkuningriigi vahelise koostööprojektina *US-UK Prize Challenge on PETs* [147], mille eesmärgiks on kokku tuua erinevate valdkondade esindajad privaatsuskaitse tehnoloogiatel põhinevate lahenduste uurimiseks, arendamiseks ja kasutuselevõtuks. Võistlust veavad USA poolt Valge Maja teadus- ja tehnoloogiaamet (*Office of Science and Technology Policy*, OSTP), USA standardimisagentuur NIST ja riiklik teadussihtasutus NSF.

Projekti alguses püstitati kaks suuremat probleemi, mille lahendustena oodatakse privaatsuskaitse tehnoloogiate kasutamist. Nendest esimene on finantskuritegude ennetamine – eesmärk on privaatsuskaitse tehnoloogiaid kasutada asutustevaheliste andmete kombineerimisel ja ühisel andmeanalüüsil, et tuvastada ja ennetada rahapesu ja muid finantskuritegusid.

Teine ülesanne on privaatsuskaitse tehnoloogiate abil üksikisikute nakatumise riski hindamine pandeemia ajal. Esimese etapi jooksul esitati võistlusele 76 tehnilist raportit, mille seast valiti võitjatena välja 12[60]. Võistluse teises etapis peavad võistlejad ehitama raportis välja pakutud lahenduse prototüübi, mille privaatsust ja turvalisust kolmandas etapis testima hakatakse.

### 5.1.2 Privaatsuskaitse tehnoloogiad õigusaktides

Koduostu järel on kõrgharidus tihti Ameerika Ühendriikide kodanike elu suurim investeering. Selleks, et tulevased tudengid saaksid kõrgema hariduse omandamisega seotud otsuseid paremini teha, on USA kongressile esitatud õigusakt *Student Right to Know Before You Go Act of 2019* [148]. Õigusakti eesmärk on koostada täpsem, täielikum ja turvalisem andmesüsteem, mille abil saab soovija kõrgharidusasutuse kohta teha päringuid.

Näiteks võib uurida kooli sisseastujate ja lõpetajate arvu või keskmist lõpetamisjärgset sissetulekut. Õigusaktis on täpsustatud, et andmete hoiustamiseks ja töötlemiseks tuleb kasutada turvalist ühisarvutust (peatükk 4.2.10) või tehnikaid, mis pakuvad sama head või paremat turvalisust ja privaatsust. Kasutatud tehnikad peavad tagama, et toorandmeid näeks ainult andmete sisestaja ja et süsteemis olevate andmete kohta ei avalikustata teistele osapooltele midagi peale õigusaktis välja toodud analüüside tulemite.

### 5.1.3 Rahvaloenduse anonüümitud ja avaandmetena avaldatud tulemite tagasituvastamine

Enne diferentsiaalprivaatsuse rakendamist, alates aastast 1990, kasutas Ameerika Ühendriikide statistikaamet rahvaloenduse andmete analüüsimisel ja tulemuste avaldamisel anonüümimist. Peamiselt kasutati andmete saalimist. Vähemal määral asendati unikaalseid väärtuseid genereeritud väärtusega, milleks 2010. aasta rahvaloendusel kasutati ka sünteetilisi andmeid [149].

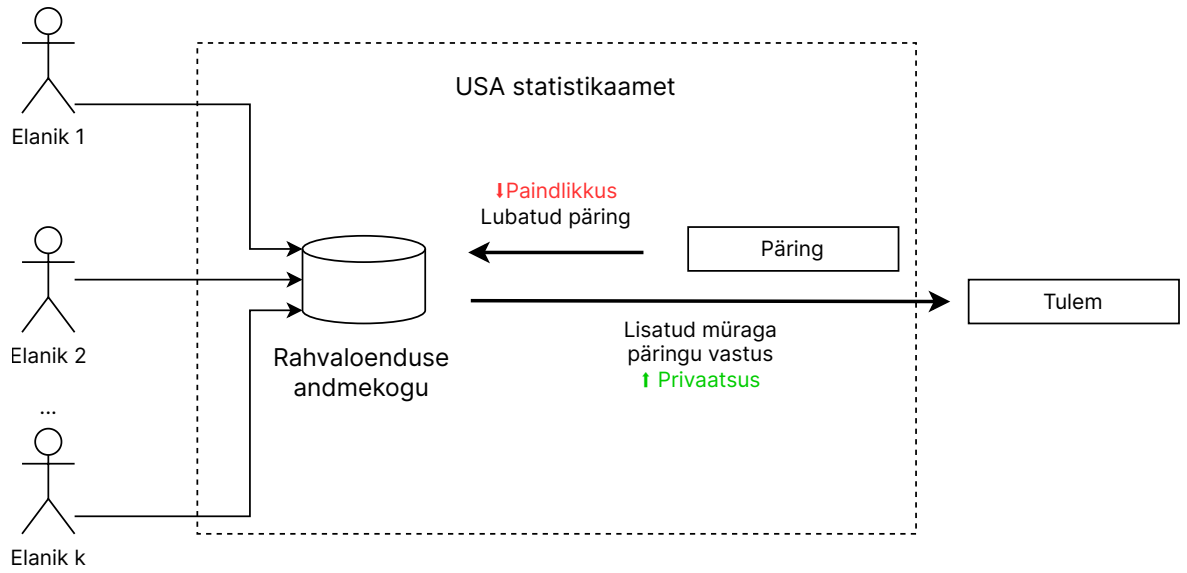
Teadlased on näidanud, et anonüümimise rakendamine ei takista tihti isikute tagasituvastamist [35]. Artiklis pakutakse välja mudel, millega saab hinnata inimese unikaalsust, ning seda kasutades näidatakse, et 99.98% ameeriklastest on võimalik tagasituvastada 15 demograafilise tunnuse põhjal. Ka USA statistikaamet suutis 2010. aasta rahvaloenduse tulemite ja kommertsandmebaasidest saadud andmete põhjal tagasituvastada 52 miljoni anonüümitud isiku geograafilise asukoha, soo, vanuse, rassilise ja etnilise kuuluvuse [150]. See viis Ameerika statistikaameti diferentsiaalprivaatsuse kasutamiseni.

[60] https://petsprizechallenges.com (viimati külastatud 02.03.2023).

# Diferentsiaalprivaatsus Ameerika Ühendriikide rahvaloendusel

| | |
|---|---|
| **Lühidalt:** Ameerika Ühendriikide statistikaamet kasutas 2020. aasta rahvaloenduse andmete analüüsimisel diferentsiaalprivaatsust, et muuta tagasituvastamise võimalikult raskeks. | **Teostamise aasta:** 2020 |
| | **Riik:** Ameerika Ühendriigid |
| | **Omanik:** Riiklik statistikaamet (US Census) |
| | **Teostaja:** Riiklik statistikaamet (US Census) |
| | **Süsteemi küpsus:** püsiv juurutus |
| **Privaatsuskaitse tehnoloogiad:**<br>1. diferentsiaalprivaatsus | **Sobivad kasutusjuhtumid:**<br>avaandmed |

**Ülevaatlik mudel:**




**Märkimisväärsed omadused:**

1. Loenduse põhjal oli aastal 2020 USA rahvaarv 331 449 281.
2. Diferentsiaalprivaatset andmete avaldamist on vaidlustatud kohtus, kuid edutult.
3. Lahendus on saanud palju meediakajastust. Üldiselt kajastatakse privaatsuse olulisust ning tuuakse eraldi välja, et väiksema privaatsusriive tõttu on küsitluste vastamisprotsent suurem, eriti vähemuste esindajate seas.
4. Kasutusjuhu positiivne ühiskondlik mõju on inimeste privaatsuse parem kaitse ja selle tõttu parem suhtumine rahvaloendusesse. Samas on tagatud piisavalt täpsed tulemused olulisemate analüüside tegemisel.

### 5.1.4 Diferentsiaalprivaatsusega rahvaloendus

Ameerika Ühendriikide statistikaamet on kogu rahvaloenduse andmete avaldamise teinud diferentsiaalprivaatsusega. 2020. aasta Ameerika Ühendriikide rahvaloendus, mille tulemusi analüüsiti ja avalikustati diferentsiaalprivaatsuse abil [149].

USA rahvaloenduse korraldamist kohustab USA põhiseadus (*Article I, Section 2*). Loendust on läbi viidud alates aastast 1790 iga kümne aasta järel. Rahvaloenduse põhjal muudetakse näiteks Ameerika Ühendriikide kongressi esindajatekoja valimisringkondi ja pannakse paika osariikide föderaalse rahastuse jaotust. Rahvaloendusel kogutavate andmete kaitset käsitlev *US Code Title* 13 keelab privaatsete andmete avaldamise või nende kasutamise muudel põhjustel kui varem täpsustatud statistiliste analüüside koostamiseks.

Rahvastikuteadlased on avaldanud arvamust, et kuigi diferentsiaalprivaatsuse kasutamine annab piisavalt täpseid tulemusi rahvaarvu kohta suuremates piirkondades, ei pruugi see nii olla väiksemates piirkondades ja rahvastiku alamgruppide osas. Probleemidena on näiteks välja toodud võimalikud ebatäpsused väiksemate piirkondade vähemuste esindatuses [151] ja imikusuremuses [152].

Süsteemi on parasjagu kajastatud ning ka kritiseeritud [153, 154, 155, 156, 157]. Diferentsiaalprivaatsuse kasutamise tõttu algatas Alabama osariik hagi, väites, et tehnoloogia kasutuselevõtu tõttu avaldatakse valimisringkondade koostamiseks vajalikud arvud liiga hilja ja avalikustatud arvud on ebatäpsed. See hagi jäeti rahuldamata [158]. Bostoni teadlaste sooritatud uuringu[159] põhjal võimaldavad uued tehnikad tasakaalustatud ja piisavalt täpset valimisringkondade jaotust.

Kaitsmaks vastajate isiklikke andmeid, on USA statistikaamet alates 1930. aastast teinud muutusi analüüside tulemuste avaldamises. Kui esialgu jäeti välja tulemid väikeste kogukondade kohta, siis 1990. aasta rahvaloendusest alates lisati kogutud andmetele ka müra. Selleks kasutati muuhulgas andmete saalimist, erindite eemaldamist või asendamist, tabelite või nende lahtrite varjamist. Vaatamata nendele abinõudele suutis USA statistikaamet 2010. aasta rahvaloenduse tulemite ja kommertsandmebaasidest saadud andmete põhjal tagasituvastada 52 miljoni isiku geograafilise asukoha, soo, vanuse, rassilise ja etnilise kuuluvuse.

2020. aasta rahvaloenduse andmete analüüsimisel kasutati diferentsiaalprivaatsuse tehnikaid, mis võimaldavad privaatsuseelarve hoolikal kasutamisel avalikustada võimalikult täpseid tulemusi, hoides samal ajal ära privaatsusriiveid. Tulemuste järeltöötluse käigus parandatati olulised loogikavead nagu negatiivne või murruline inimeste arv teatud piirkonnas. Erandina avalikustati täpse väärtusena üksikud olulised statistikud nagu iga osariigi elanike arv.

### 5.1.5 Turvalise ühisarvutuse kasutamine palgalõhe uurimisel

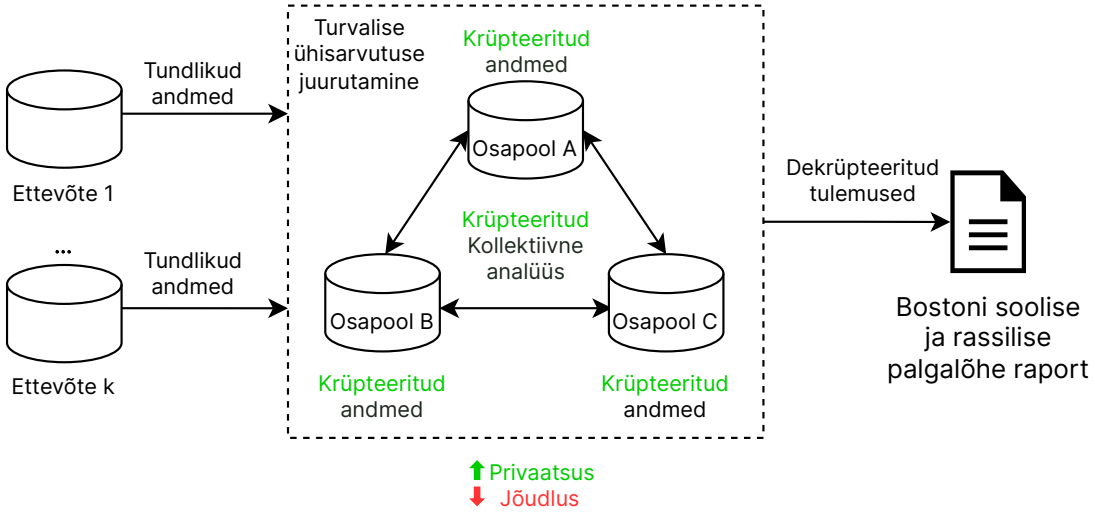
Bostoni linna naiste tööjõu nõukogu (*Boston Women Workforce Council*, BWWC) viib iga kahe aasta järel läbi Bostoni soolise ja rassilise palgalõhe uuringut, et mõõta ja analüüsida sealsete ettevõtete, ametkondade ja muude organisatsioonide tööliste andmeid. Tundlikkuse tõttu analüüsitakse andmeid turvalise ühisarvutuse abil [160, 161, 162].

Turvaline ühisarvutus võimaldab uurida palgaandmeid vastavalt inimeste soole, rassile, töökategooriale ilma privaatseid andmeid avaldamata. Uuring näitas, et sotsiaalseid probleeme saab tehnoloogia abil analüüsida ilma inimeste privaatsust riivamata [163], mistõttu soovitatakse privaatsuskaitse tehnoloogiatel põhinevaid lahendusi ka muude sotsiaalsete probleemide uurimiseks [164].

# Bostoni palgalõheuuring

| | |
|---|---|
| **Lühidalt:** Bostoni linnas kogutakse regulaarselt andmeid, et hinnata erinevate inimgruppide palgatasemete võrdsust. | **Teostamise aasta:** 2015, 2016, 2018, 2020 |
| | **Riik:** Ameerika Ühendriigid (*Greater Boston Area*) |
| | **Omanik:** *Boston Women's Workforce Council* (BWWC) |
| | **Teostaja:** BWWC üheskoos Bostoni ülikooli Hariri instituudiga |
| | **Süsteemi küpsus:** püsiv juurutus |
| **Privaatsuskaitse tehnoloogiad:**<br>1. turvaline ühisarvutus | **Sobivad kasutusjuhtumid:**<br>privaatne analüütika |

**Ülevaatlik mudel:**




**Märkimisväärsed omadused:**

1. Aastal 2020 koguti andmeid 134 tööandjalt enam kui 156 000 töötaja kohta, kelle kollektiivne aastapalk ületas 17.4 miljardit dollarit.
2. Bostoni palgalõheuuring oli tehnoloogiat ja valdkonda silmas pidades esimene omalaadne Ameerika Ühendriikides. See oli esimene kord, kui palganumbrite aruandlus toimus anonüümsuse ja vabatahtlikkuse alusel.
3. Uuring näitas, et sotsiaalseid probleeme saab tehnoloogia abil analüüsida ilma inimeste privaatsust riivamata ning selline analüüsiviis on nüüd muutunud regulaarseks.

### 5.1.6 Ameerika Ühendriikide tehnoloogiafirmade aktiivsus PETide rakendamisel

Tehnoloogiafirma Google on esimene ja üks suurimaid liitõppe juurutajaid. Aastal 2017 tutvustasid nad nii tehnoloogiat kui selle esimest kasutusjuhtu: liitõpe nutitelefonide virtuaalklaviatuuri Gboard tekstiprognoosi ja sõnade automaatjätkamise jaoks [50]. Liitõppe juurutamise lihtsustamiseks on Google abiga välja töötatud vabavaraline teek *Tensorflow Federated*[61].

COVID-19 pandeemia alguses kasutas Google diferentsiaalprivaatsust, et koostada ja avaldada privaatsust säilitavaid raporteid inimeste liikumismustrite kohta. Selleks jälgiti nutiseadmeid, millel oli sisse lülitatud asukoha ajaloo säilitamine. Loeti kokku, mitu erinevat sellise nutiseadme kasutajat külastas teatud avalikku kohta. Tulemused agregeeriti ja anonüümiti, seejärel rakendati diferentsiaalprivaatsust. Koostatud raportite eesmärgiks oli informeerida terviseameteid ja aidata neil võtta vastu otsuseid nakatumiste vähendamiseks [165].

Google on välja arendanud ka vabavaralise täishomomorfse krüpteerimise teegi [166].

Meta arendab ja testib privaatsuskaitse tehnoloogiaid peamiselt isikustatud reklaamide pakkumiseks ja nende tõhususe hindamiseks [167]. Näiteks võib turvaline ühisarvutus tulevikus võimaldada Meta platvormidel reklaamijatel hinnata suunatud reklaamide tõhusust, s.t. arvutada, kui suur osa reklaamile vajutajatest sooritab uuele lehele suunatult sealt ostu[62]. Lisaks arendab Meta liitõppesüsteeme, mis kasutavad tulemuste agregeerimisel diferentsiaalprivaatsust ja usaldatud käivituskeskkondi. [168].

Apple kasutab liitõpet masinõppe mudelite treenimiseks [169]. Tõenäoliselt kõige tuntum kasutuslugu on isikupõhine kõnetuvastus, nimelt suudab Siri õppida isiku sõnavara ja keelekasutust liitõppe abil. Lisaks kasutatakse liitõppega treenitud mudeleid näiteks uudiste soovitamiseks, tekstiprognoosiks ja *emoji*'de soovitamiseks, ning jõudlusprobleeme tekitavate andmetüüpide tuvastamiseks. Andmete agregeerimisel kasutatakse diferentsiaalprivaatsust ja pseudonüümimist.

Safari kasutab teatud juhtudel analüütiliste andmete korjamisel diferentsiaalprivaatsust.[63] Kaardirakendus Apple Maps kasutab ühe isiku seadmete sünkroonis hoidmiseks otspunktkrüpteerimist ning ajutisi juhuslikke identifikaatoreid ning asukoha hägustamist. Rakenduse laiendite jaoks kasutatakse aedikuid [170].

Suhtlusrakendused FaceTime ja iMessage võimaldavad otspunktkrüpteeritud suhtlust [170].

Microsoft on aktiivne privaatsuskaitse tehnoloogiate arendamisel ja kasutuselevõtmisel. Nende välja töötatud karkass CCF võimaldab turvaliste käivituskeskkondade abil plokiahela võrgus krüpteeritud tehinguid sooritada, pakkudes tehingute tervikluse kontrollimise võimalust koos paremate konfidentsiaalsusgarantiidega. CCF on mastabeeritav, võimaldades töödelda üle 50 000 tehingu sekundis, ning suudab toetada olemasolevaid plokiahelprotokolle. [171].

Microsofti platvorm EXP toob töötaja mugavuseks kokku erinevad kommunikatsiooni, õppimise ja analüütika tööriistad. Platvormi alamteenus Viva Insights kasutab umbisikustamist ja diferentsiaalprivaatsust, et võimaldada organisatsiooni näitajate uurimist, riivamata töötajate privaatsust [172].

Microsofti veebibrauseri Edge parooligeneraator ja -monitor kasutavad homomorfset krüpteerimist, et sooritada päringuid krüpteeritud andmete põhjal, nägemata pääsumandaate. Paroolimonitor teavitab kasutajat, kui tema paroolihalduri andmefailist on üks või mitu parooli lekkinud [173].

Suhtlusrakendus Signal on välja töötatud Ameerika Ühendriikides. Algselt ettevõtte Whisper Systems ning seejärel ettevõtte Open Whisper Systems poolt välja arendatud rakenduste RedPhone ja TextSecure järeltulijana on Signal tuntud vestluste otspunktkrüpteerimise ja sellega kaasnevate privaatsusgarantiide poolest. Moxie Marlinspike, kes oli nii Whisper Systems kui Open Whisper Systems asutaja, pani koos WhatsAppi kaasasutaja Brian Actoniga aastal 2018 aluse mittetulunduslikule organisatsioonile Signal Technology Foundation, mis sellest ajast suhtlusrakendust edasi arendab [124]. Praegu kasutavad

[61] Tensorflow Federated https://www.tensorflow.org/federated (viimati külastatud 02.03.2023).
[62] https://github.com/facebookresearch/fbpcf (viimati külastatud 02.03.2023).
[63] Safari Privacy Overview https://www.apple.com/safari/docs/Safari_White_Paper_Nov_2019.pdf (viimati külastatud 02.03.2023).

tehnoloogia aluseks olevat protokolli salajasteks vestlusteks ka kinnise lähtekoodiga suhtlusrakendused nagu Google Messages, Facebook Messenger ja Skype.

### 5.1.7 Teadus- ja standardimisprogrammid ning nende tulemid

Sibulmarsruutimise tehnoloogia töötati välja USA mereväe teaduslaboris möödunud sajandi lõpus, seejärel arendas seda edasi USA Kaitseministeeriumi teadusagentuur DARPA. Aastal 1998 patenteeris tehnoloogia merevägi. Aastal 2002 pandi alus Tor projektile ja aastal 2006 mittetulunduslikule organisatsioonile The Tor Project, Inc, mida veavad vabatahtlikud.[64][65]

Ameerika teadusagentuurid nagu DARPA, IARPA, NSF jt rahastavad süsteemselt privaatsuskaitse tehnoloogiaid arendavaid programme nagu PROCEED[66], Brandeis[67], SIEVE[68], HECTOR[69].

Standardimisagentuur NIST on USA Kaubandusministeeriumi mitteregulatiivne agentuur, mis arendab metroloogiat, standardeid ja tehnoloogiaid. NIST on avaldanud diferentsiaalprivaatsuse kasutusjuhtudele ja tehnilistele aspektidele pühendatud blogipostituste sarja[70]. Sarja eesmärk on anda talitusprotsesside omanikele ülevaade diferentsiaalprivaatsuse põhitõdedest ning aidata privaatsustehnikutel diferentsiaalprivaatsusega seotud tööriistu[71] implementeerida. Ekspertidest vabatahtlike abiga loodetakse blogisarjast teha põhjalikum ja laiaulatuslikum diferentsiaalprivaatsusjuhend. Lisaks diferentsiaalprivaatsusele on blogisarjas uuritud ka privaatsust säilitavat krüptograafiat, ohumudeleid ning andmebaaside päringute loomist.

NIST algatas koostöös USA ja Ühendkuningriigi asutuste ja ametkondadega privaatsuskaitse tehnoloogiatega seotud innovatsioonile pühendatud võistluse *US-UK Prize Challenges on Privacy Enhancing Technologies* [147].

Eelmainitud algatused on osa asutuse riskihalduse raamistiku teekaardist, mille eesmärk on toetada pidevat arengut ja koostööd avaliku sektori ja erasektori asutuste privaatsusriskide haldamisel [174].

## 5.2 Holland

### 5.2.1 Privaatsuskaitse tehnoloogiad riigi teekaardil

Hollandi Majandus- ja Kliimapoliitika ministeeriumi eestvedamisel koostatakse krüptograafia teekaarti, milles on osa ka privaatsuskaitse tehnoloogiatel [175]. Eraldi tuuakse dokumendis välja võimalusi privaatsuskaitse tehnoloogiate kasutamiseks tervishoiusektoris, kus andmed on eriliigilised, aga samas väärtuslikud haiguste uurimisel.

### 5.2.2 Teadus- ja standardimisprogrammid

Rakendusuuringute organisatsioon TNO uurib teadusprogrammi ERP raames, kuidas rakendada krüptograafiat ühiskondlike probleemide lahendamiseks. Muuhulgas kavatsetakse uurida privaatsuskaitse

[64] https://www.torproject.org (viimati külastatud 02.03.2023).
[65] https://en.wikipedia.org/wiki/Tor_(network) (viimati külastatud 02.03.2023).
[66] PROgramming Computations on EncryptEd Data (PROCEED) https://www.darpa.mil/program/programming-computation-on-encrypted-data (viimati külastatud 24.02.2023).
[67] Brandeis https://www.darpa.mil/program/brandeis (viimati külastatud 24.02.2023).
[68] Securing Information for Encrypted Verification and Evaluation (SIEVE)https://www.darpa.mil/program/securing-information-for-encrypted-verification-and-evaluation (viimati külastatud 24.02.2023).
[69] Homomorphic Encryption Computing Techniques With Overhead Reduction (HECTOR) https://www.iarpa.gov/research-programs/hector (viimati külastatud 24.02.2023).
[70] Differential Privacy Blog Series https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space/focus-areas/de-id/dp-blog (viimati külastatud 02.03.2023).
[71] De-identification Tools https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space/focus-areas/de-id/tools (viimati külastatud 02.03.2023).

tehnoloogiate kasutamist pilditöötlusel videovalvesüsteemides, logistiliste ahelate optimeerimisel ja hajusa targa elektrivõrgu loomisel [176].

### 5.2.3 Privaatsuskaitse tehnoloogiad statistikas

Hollandi statistikaamet CBS on privaatsuskaitse tehnoloogiaid uurinud juba alates 2018. aastast. Tehnoloogiate abil saaks statistilistel analüüsidel kaasata andmeid osapooltelt, kes andmeid muidu jagada ei soovi või näiteks juriidilistel põhjustel neid jagada ei tohi. Hetkel uurib statistikaamet koostöös Groningeni ja Maastrichti ülikoolidega võimalusi üksteise ühendandmete analüüsimiseks ilma andmeid jagamata. Tulevikus võivad privaatsuskaitse tehnoloogiad aidata statistikaametil hoiustada andmeid turvalisemalt ja jätkusuutlikumalt [177].

2023. aasta alguses avaldas CBS, et nad uurivad ja katsetavad sünteetiliste andmete genereerimist privaatsuskaitse tehnoloogiana. Kogemuste saamiseks sünteesisid nad Hollandi iduettevõtte Syntho valmistatud tarkvarapaketi abil Hollandi äriregistri andmeid.[72] Statistikaameti hinnangul oli andmete genereerimine edukas. Samas leiti, et genereeritud andmeid peaks kasutama peamiselt asutusesiseselt tarkvara testimiseks. Lisaks toodi välja, et sünteetiliste andmete avaldamisel võib lekkida infot algse andmestiku kohta ning vastavaid riske peab rohkem uurima.

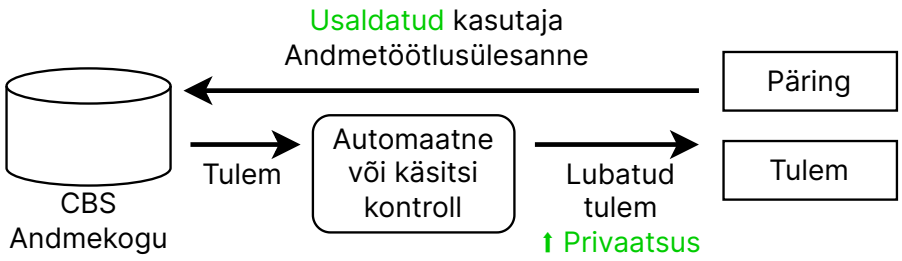
Hollandi statistikaamet CBS lubab kasutada pseudonüümitud mikroandmeid välistel osapooltel turvalises kaugpääsu keskkonnas [178]. Süsteem on püsivas arenduses.

[72] https://www.cbs.nl/en-gb/about-us/innovation/project/what-is-synthetic-data- (viimati külastatud 02.03.2023).

# Hollandi statistikaameti analüütiku töökoht

| | |
|---|---|
| **Lühidalt:** Kasutaja võib omalt poolt keskkonda andmeid üleslaadida ja viia läbi statistilisi analüüse või treenida masinõppe mudeleid. | **Teostamise aasta:** Alates 2006 |
| | **Riik:** Holland |
| | **Omanik:** Riiklik statistikaamet |
| | **Teostaja:** Riiklik statistikaamet |
| | **Süsteemi küpsus:** püsiv juurutus |
| **Privaatsuskaitse tehnoloogiad:**<br>1. pseudonüümimine<br>2. juurdepääsupiirangu rakendamine | **Sobivad kasutusjuhtumid:**<br>avaandmed |

**Ülevaatlik mudel:**




**Märkimisväärsed omadused:**

1. Juurdepääs andmetele on piiratud nii juriidiliste (konfidentsiaalsuslepingud), kui ka tehniliste meetmetega (VPN ja Citrix kliendirakendus). Lisaks peab kasutajal olema pääsmik, PIN-kood, mobiiltelefon ja kasutajanimi/parool. Kasutaja ei saa tulemusi keskkonnast välja võtta ilma statistikaameti nõusolekuta. Enne tulemuste väljastamist hinnatakse nende tundlikkust, et vältida nii mikroandmete kui võimalike privaatsust riivavate tulemuste lekkimist.
2. Süsteemi on hinnanud komitee, mis tõi välja valupunkte ja juhiseid turvalisuse parandamiseks. Tõsiseid puudujääke ei tuvastanud, kuid alates 2021. aasta augustist on komitee koostatud raporti põhjal sisse viidud uuendusi.
3. Kuigi tõsiseid suuremahulisi ründeid süsteemi vastu pole teada, on kasutajad ise tunnistanud, et huvi korral oleks võimalik pseudonüümitud andmetest üksikisikuid tuvastada.

### 5.2.4 Privaatsuskaitse tehnoloogiate pilootprojektid tervishoiusektoris

Hollandis on tervishoiusektoris läbi viidud mitmeid pilootprojekte. Personaalmeditsiini projekti CARRIER eesmärk on isiku südame isheemiatõve ennetamine, prognoosimine ja haigestumisriski hindamine isiku elustiili ning meditsiiniliste ja muude isiklike andmete põhjal. Mitme koostööpartneri, sealhulgas Hollandi ülikoolide, statistikaameti, kliiniliste asutuste andmete privaatne, eetiline ja seaduslik teisene kasutamine nõuab tugevat juriidilist ja tehnilist alust. Projektis CARRIER plaanitakse andmelinkimiseks ja masinõppeks kasutada turvalist ühisarvutust, homomorfset krüptograafiat ja liitõpet [8].

Statistikaamet CBS, Zyderlandi haigla, tervisekindlustusfirma CZ ning Hollandi rakendusuuringute organisatsioon TNO viisid koos läbi personaalmeditsiini pilootprojekti, mille eesmärgiks oli hinnata soole ärritussündroomi (IBS) patsientidele mõeldud telefonirakenduse kasulikkust. Koostatud hinnangu põhjal saab otsustada, kas kindlustusfirma peaks selle rakenduse kasutamisega seotud kulud patsiendile hüvitama.

Projekti käigus kasutati privaatset hulkade ühisosa meetodit (alamosa turvalisest ühisarvutusest), homomorfset krüpteerimist ja turvalist ühisarvutust, et ühendada CBS, CZ ja Zyderlandi haigla andmed ning analüüsida neid, kaitstes samaaegselt nende privaatsust. Tulemuste ja järeldustena tõid asutused välja palju juriidilisi raskusi. Andmete käsitluses on oluline kooskõlastus üldiste andmekaitsemäärustega nagu IKÜM, aga ka spetsiifilisemate meditsiiniandmetega seotud seadustega. Lisaks on oluline omada korrektseid koostöölepinguid ning küsida kirjalikult luba uuringus osalevate inimestelt nende andmete teiseseks kasutamiseks. Funktsionaalse poole pealt rõhutati järgmiste aspektide olulisust: hea kvaliteediga metaandmete ja sünteetiliste andmete olemasolu, täpsed linkimisidentifikaatorid erinevates andmebaasides ning statistiline paljastustõrje oht teiste lahenduste puhul [179].

### 5.2.5 Energiatarbimise prognoosimine ja tasakaalustamine

Elektrivõrguoperaatorid peavad oskama prognoosida ja tasakaalustada elektrivõrgu nõudlust ja tootlust. Jätkusuutliku energia lokaalne tootmine on teinud selle keeruliseks, mistõttu Hollandi elektrivõrguoperaator Stedin otsib võimalusi privaatsust kaitsvalt mõõta kasutajate energiatarbimist ja -tootlust ning prognoosida nende põhjal energiavajadusi ja ennetada võrgu ülekoormust. Selleks viib Stedin koos ettevõtetega Technolution Spark ja Roseman Labs läbi pilootprojekti, milles otsitakse võimalusi rakendada turvalist ühisarvutust majapidamiste energiatarbmise analüüsimiseks ilma klientide privaatsust riivamata. Juba valminud prototüüp agregeeris turvalise ühisarvutuse abil kuuest kaugarvestist kogutud andmed ning kuvas seejärel tulemused võrguoperaatori Stedin innovatsioonilaboris asuvale koondpaneelile [180].

## 5.3 Jaapan

### 5.3.1 Privaatsuskaitse tehnoloogiate arendajate liit

2022. aasta augustis pandi Jaapanis alus privaatsuskaitse tehnoloogiate vabatahtlike ühendusele *Privacy Tech Association*[73]. Ühenduse eesmärk on juurutada privaatsuskaitse tehnoloogiate kasutamist ettevõtetes, mis töötavad andmetega. Muuhulgas uuritakse anonüümimist, turvalist ühisarvutust, diferentsiaalprivaatsust, andmesünteesi. Ühendusega on liitunud mitmed tehnoloogia-, juura- ja andmeteaduseksperdid akadeemilistest asutustest ja erasektorist.

### 5.3.2 Andmekaitseõiguse ja privaatsuskaitse tehnoloogiate alane teavitustöö

Jaapani õigusteaduse ja infosüsteemide instituut JILIS on aastal 2016 asutatud organisatsioon, mis keskendub teadustegevusele ning infosüsteemidega seotud uutele poliitikatele ja seadustele. Muuhulgas avaldab instituut andmekaitsega seotud soovitusi ja arvamusi, tuues esile puudujääke või ebatäpsusi andmete hoiustamise ja jagamisega seotud dokumentides, sh infosüsteemide spetsifikatsioonides, raportites, seadustes ja seaduseelnõudes. Tihti rõhutavad nad vajadust eristada selgelt anonüümimist,

[73] https://privacytech-assoc.org (viimati külastatud 02.03.2023).

pseudonüümimist ja krüpteeritud andmeid ning toovad esile iga tehnikaga seotud puudusi ja võimalusi[74].

## 5.4 Kanada

### 5.4.1 Kanada on lõimitud privaatsuse kodumaa

Lõimitud privaatsuse (*Privacy-by-Design*) konseptsioon arendati välja Kanadas 1990-ndatel tollase Ontario Informatsiooni- ja privaatsusvoliniku Ann Cavoukiani eestvedamisel. Lõimprivaatsetes süsteemides arvestatakse privaatsusnõuetega süsteemi kogu elutsükli jooksul [181].

Ann Cavoukian oli Ontario informatsiooni- ja privaatsusvolinik aastatel 1997-2014. Tema eestvedamisel viidi läbi mitmeid uuringuid ja toodi esile privaatsusriive probleeme suurprojektides. Aastal 2007 teatas Toronto transpordiamet TTC kavatsusest laiendada ühistranspordisüsteemide videovalvesüsteemi. *Privacy International* avaldas muret vastava lahenduse privaatsuse osas ning kahtles lahenduse kasulikkuses.

Ann Cavoukiani eestvedamisel koostati raport, milles toodi esile potentsiaalseid murekohti ja esitleti ka võimalusi privaatsuskaitse tehnoloogiate kasutamiseks videovalve analüüsimiseks. Täpsemalt pakuti välja lahendus, kus videosse jääv isikutuvastusteave krüpteeritakse ning dekrüpteerimiseks vajalikud võtmed oleks ainult selleks volitatud osapooltel [182].

### 5.4.2 Andmekaitseagentuuri tugev roll

Kanadas andmekaitse alast nõu ja informatsiooni jagav amet *Office of the Privacy Commissioner of Canada* avaldas 2017. aasta ülevaatliku raporti privaatsuskaitse tehnoloogiatest [183]. Raportis klassifitseeritakse tehnoloogiad nende kasutusviiside ja -võimaluste järgi ning antakse lühike ülevaade erinevate tehnoloogiate enamlevinud kasutusjuhtudest ja tuntumatest juurutajatest. Lisaks arutatakse, miks pole PETid rohkem levinud ja kuidas neid populariseerida.

### 5.4.3 Tugev kontroll murujuuretasandil

2017. aastal avaldas Alphabeti tütarfirma Sidewalk Labs plaani arendada Torontos peaaegu viie hektari suurune Quayside linnaosa, kus töötataks välja ja testitaks uusi andmepõhiseid linnaplaneerimise tehnoloogiaid. Kriitikud leidsid, et Google emafirmale Alphabetile kuuluva Sidewalk Labsi juurutatud süsteem loob liiga palju privaatsusriske. Projekt leidis palju meediakajastust ja hakkas andmete kogumise hirmude valguses venima. Projekt tühistati 2020. aastal. Sidewalk Labs tõi tühistamise põhjusena välja koroonapandeemiaga seotud majandusliku ebakindluse [184].

### 5.4.4 Sünteetilised teisikud terviseandmetest

Kanada ettevõte Replica Analytics on välja arendanud tarkvara meditsiiniandmete sünteesimise teenuse pakkumiseks. Sünteesitud andmete genereerimiseks kasutatakse masinõppe ja tehisnärvivõrkude meetodeid. Lisaks andmete genereerimisele väljastatakse raport, mis kirjeldab sünteesitud andmete kasulikkust - näiteks seda, kui hästi säilivad andmete statistilised omadused nagu tunnustevaheline korrelatsioon.

Teenuse abil saavad sünteesitud andmeid kasutada ametnikud, teadlased, analüütikud, kes muidu tundlikele andmetele ligi ei pääseks, muuhulgas kasutati tarkvara Ontario COVID-19 patsientide andmebaasi sünteetilise versiooni loomiseks. Tehnoloogiat on kasutatud ka jämesoolevähi andmestiku sünteesimiseks, sealjuures viidi tehnoloogia valideerimiseks läbi samu statistilisi analüüse nii sünteetilise kui pärisandmestiku peal ning võrreldi nende tulemusi. Uuring näitas, et analüüsid annavad mõlema andmestiku peal sarnaseid tulemusi. Replica Analytics uurib lisaks sünteetiliste andmete kasulikkuse tunnusomadustele ka nende privaatsusriske, lisaks viib ettevõte läbi seminare ja on avaldanud raamatuid sünteetiliste

[74] https://www.jilis.org/proposal (viimati külastatud 02.03.2023).

andmete genereerimisest[75].

### 5.4.5 Privaatsuskaitse tehnoloogiad riiklikus statistikas

Kanada statistikaamet Statistics Canada on kasutanud kontseptsiooni tõenduses privaatset hulkade ühisosa meetodit (alamosa turvalisest ühisarvutusest) koos turvalise ühisarvutusega. Testitud lahendus võimaldaks hinnata kolmanda osapoole andmebaasi kattuvust populatsiooniga või statistikaagentuuri andmebaasiga, jagamata andmeid. Andmete kattuvus annab aimu, kas andmete ühendamine edasisteks analüüsideks tasuks ennast ära.

Kanada statistikaamet testis masinõppe mudelite treenimist sünteesitud andmete peal, kasutades homomorfse krüpteerimise meetodeid. Testülesandeks oli mitme jaemüüja andmete peal treenida tehisnärvivõrk, mis liigitaks tootekirjeldusi. Tulemusena leiti, et päriselulise probleemi lahendamiseks on võimalik mõistliku aja jooksul treenida tehnisnärvivõrk, kasutades homomorfset krüptograafiat. Väljatöötatud lahendust saaks tulevikus kasutada pilves asuva mudeli treenimiseks ja kasutamiseks nii, et erinevad osapooled saadavad andmeid pilve vaid krüpteeritud kujul.

Häkatonide jaoks on tihti vaja pärisandmetele statistiliselt lähedasi andmeid. Kanada statistikaamet genereeris aastatel 2018 ja 2019 häkatonide jaoks sünteetilisi andmeid pärisandmete põhjal. Selleks kasutati andmete imputeerimise meetodit. Statistikaameti hinnangul oli andmesüntees edukas, sest võimaldas ametil saada kogemusi sünteetiliste andmete genereerimisega ning mõlemad häkatonid viidi läbi sünteesitud andmetega [8].

## 5.5 Prantsusmaa

### 5.5.1 CNIL toetav roll privaatsuskaitse tehnoloogiate juurutamisel

1978. aastal asutatud Prantsusmaa andmekaitse agentuur CNIL on välja töötanud ja avaldanud käsitlusi ja juhendmaterjale mitmetest privaatsuskaitse tehnoloogiatest. 2014. aastal avaldas CNIL käsitluse pseudonüümimis- ja anonüümimistehnikatest [185]. Dokumendis antakse ülevaade andmete saalimisest, k-anonüümimisest, l-hajutatusest, t-lähedusest, müra lisamisest ja diferentsiaalprivaatsusest ning selgitatakse nende tehnikate põhimõtteid, tuuakse välja nendega seotud tugevusi ja nõrkusi ja nende rakendamisel tehtavaid tüüpvigu.

CNIL korraldab regulaarselt teaduskonverentsi *Privacy Research Day*, mille teemade hulgas on privaatsust säilitavad tehnoloogiad.[76] Üritusele oodatakse ettekannetena päriselulisi näiteid privaatsuskaitse tehnoloogiate rakendamisest.

Koos Prantsusmaa arvutiteaduse instituudiga Inria annab CNIL juba seitsmendat korda välja privaatsuskaitse auhinda *CNIL-Inria Privacy Award*, mille eesmärk on edendada privaatsus- ja andmekaitsealast teadustööd, keskendudes privaatsuslõimele, algoritmide läbipaistvusele, privaatsuskaitse tehnoloogiate juurutamisele, anonüümimisele ja privaatsusriskide analüüsile.[77]

### 5.5.2 Prantsusmaa suveräänne tehnoloogiaarendus

Prantsusmaa tugev soov digitaalse suveräänsuse järele on suunanud neid mitmeid tehnoloogiad uuesti arendama ja välja töötama, sh ehitasid nad oma unikaalse COVID-19 lähikontaktide tuvastamise rakenduse StopCovid. Erinevalt DP-3T tehnoloogial põhinevast lahendustest, kasutas StopCovid tsentraliseeritud lähenemist. Puuduvate privaatsusgarantiide tõttu pälvis rakendus palju kriitikat ning Apple ei võimaldanud seda enda nutiseadmetel kasutada. Rakenduse vastu oli prantslastel võrdlemisi väike huvi, selles süüdistas valitsus aga ebapiisavat teavitustööd [186].

[75] https://replica-analytics.com/knowledge-base (viimati külastatud 02.03.2023).

[76] CNIL, Call for Papers: Privacy Research Day 2023 https://www.cnil.fr/en/call-papers-privacy-research-day-2023 (viimati külastatud 02.03.2023).

[77] Launch of 7th edition of CNIL-Inria Privacy Award https://www.cnil.fr/en/launch-7th-edition-cnil-inria-privacy-award (viimati külastatud 02.03.2023)

### 5.5.3 Liitõppe rakendamine terviseuuringutel

Prantsuse-USA biotehnoloogia ettevõte Owkin arendab avatud lähtekoodiga liitõppe platformi Substra, mis võimaldab terviseasutustel teha koostööd uute ravimite väljaarendamisel ning personaalmeditsiini alastel uuringutel. Platvormi abil treeniti nelja prantsuse haigla andmete peal masinõppe mudel, mis prognoosib neoadjuvantse keemiaravi mõju kolmekordse negatiivse rinnavähiga patsientidele [187].

MELLODDY projektis, millest võttis osa kümme terviseasutust, kasutati Substra platvormi ravimite avastamiseks. Tegu oli ühe suurima farmaatsiaalase tehisõppe koostööprojektiga [188]. 2021. aastal moodustatud konsortsium ScanCovIA kasutas Substra platvormi, et treenida tehisnärvivõrke COVID-19 patsiendi kopsutomograafia põhjal haiguse kulu [189].

Owkin on partner ka 2022. aastal alanud OncoLab projektis, mille eesmärgiks on teha terviseasutuste andmed privaatsust kaitsvalt kättesaadavaks kõigile projekti liikmetele [190]. OncoLab Projekti juhib ettevõte Arkhn, mis keskendub samuti privaatsuskaitse tehnoloogiatele tervishoiusektoris. Täpsemalt kasutavad nad liitõppe ja diferentsiaalprivaatsuse lahendusi.[78]

## 5.6 Singapur

### 5.6.1 Privaatsuskaitse tehnoloogiate liivakast

Märkimaks kümne aasta möödumist andmekaitseseaduse PDPA (*Personal Data Protection Act*) vastuvõtmisest Singapuris, seadis sealne valitsusagentuur IMDA 2022. aasta suvel üles turvalise testimiskeskkonna privaatsuskaitse tehnoloogiatega seotud pilootprojektideks [191]. Projektide eesmärgiks on tuvastada, millised tehnoloogiad aitaksid ettevõtteid andmete jagamise probleemide puhul ja millised on erinevate tehnoloogiate piirangud. Pilootprojektide põhjal kavatsevad IMDA ja andmekaitseagentuur PDPC tuvastada tarkvarad, mida kasutatakse privaatsuskaitse tehnoloogiate rakendamisel ja arendada välja standardeid ning poliitikaid tehnoloogiate kasutuselevõtuks.

### 5.6.2 Innovatsioon COVID-19 lähikontaktide tuvastamisel

Singapuri *TraceTogether* rakendus aitab *Bluetooth* tehnoloogia abil tuvastada COVID-19 lähikontakte [192]. Rakenduse kasutajad saadavad kokkupuute korral omavahel ajutisi juhuslikke võtmeid. Kui üks osapool peaks haigestuma ja selle rakenduses vastavalt ära märgib, saab teine teavituse, et on olnud nakatanuga lähikontaktis. Vältimaks teavitusi lühiajaliste kokkupuudete või läbi seina naabritega vahetatud võtmete tõttu, hinnatakse nakatumise kontrollimisel kasutajate lähikontaktis viibimise aega ja selle aja vältel olnud signaali tugevust. Kuigi rakendus pakub privaatsusgarantiisid teiste kasutajate eest, on selle aluseks olevat *BlueTrace* tehnoloogiat kritiseeritud, sest ajutisi võtmeid jagatakse välja tsentraalselt. Seega peab nakatunu üles laadima enda kontaktide logi terviseteenusepakkuja serverisse, kus võtmed vastavate kasutajatega lingitakse. Lähikontaktsete teavitamiseks võetakse nendega ühendust vastavast asutusest.

*TraceTogether*'iga sarnase kuid detsentraliseeritud võtmetega versiooni lähikontaktide tuvastamise süsteemi arendasid välja ning viisid juurutusteni Euroopa teadlased [193].

## 5.7 Ühendkuningriik

### 5.7.1 Ameerika Ühendriikide ja Ühendkuningriigi PET programm

Ühendkuningriik ja Ameerika Ühendriigid on üheskoos algatanud privaatsuskaitse tehnoloogiate programmi ja võistluse *US-UK Prize Challenges on Privacy Enhancing Technologies* [194]. Ühendkuningriigi poolt veab võistluse korraldamist ja planeerimist andmeeetika- ja innovatsioonikeskus CDEI (*Centre for Data Ethics and Innovation*).

[78] https://arkhn.org/our-technologies (viimati külastatud 02.03.2023).

### 5.7.2 Teadus- ja standardimisprogrammid

Ühendkuningriigi riiklik teaduste akadeemia *Royal Society* andis aastal 2023 välja privaatsuskaitse tehnoloogiate aruande [6]. Aruande eesmärk on anda ülevaade kõige lootustandvamatest privaatsuskaitse tehnoloogiatest ja nende rakendustest, propageerida tehnoloogiate kasutuselevõttu Ühendkuningriigis, tutvustada erinevate tehnoloogiate kasutusvõimalusi ja ajendada nende standardiseerimist.

Raportis tuuakse välja, et teadmised tehnoloogiate olemasolust ja olemusest on vähelevinud ning uute väheuuritud tehnoloogiate kasutuselevõtt tundub paljudele ettevõtetele pigem liiga riskantne. Peamisteks valukohtadeks privaatsuskaitse tehnoloogiate juurutamisel on standardite, tuntud kasutusjuhtude ja firmasiseste ekspertide puudumine. Uute tehnoloogiate vastu puudub usaldus ning nende kasutuselevõtt tundub keeruline ja ressursse nõudev. Lahenduseks võiks olla standardite ja juhendite koostamine, tihedam koostöö teadurite ja potentsiaalsete juurutajate vahel ning riiklik strateegiline lähenemine tehnoloogiate kasutuselevõtmisel.

Aruanne toob näiteid privaatsuskaitse tehnoloogiate kasutamisest, keskendudes sellele, et potentsiaalsed kasutajad teaksid, mida tehnoloogiad teha võimaldavad, ja milliseid probleeme nende kasutamisega vältida saab. Raportis on välja toodud võimalikke ja tegelikke kasutusjuhte tervishoiu-, energia-, sotsiaal-, finantsvaldkonnast. *Royal Society* uuris ka avaliku sektori huvi privaatsuskaitse tehnoloogiate vastu. Järeldusena toodi välja, et peamiselt on küsitletud asutused huvitatud pseudonüümimisest, anonüümimisest, andmesünteesist, diferentsiaalprivaatsusest ja liitõppest.

### 5.7.3 Liitõppe ja diferentsiaalprivaatsuse prototüüpimine

Ühendkuningriigi õigussektori kaasajastamisele pühendatud algatuse LawtechUK eestvedamisel loodi prototüüp, mis võimaldab juriidilistel asutustel saada teadmusi nende ühendandmetelt, jagamata lokaalseid andmeid [195]. Prototüüp võimaldas privaatsust säilitavalt analüüsida vastutusklausleid asutuste *Norton Rose Fulbright*, Vodafone, Ashurst ja *Solicitors Regulatory Authority* pilvtarkvara dokumentidest ja koostas nende põhjal raporti. Kokku analüüsisid asutused 46 doc.x ja pdf formaadis dokumenti, millest algoritm suutis edukalt analüüsida 87%. Prototüübi loomisel kasutati *RegulAItion AIR Platformi*. Privaatsuskaitse tehnoloogiatest kasutati liitõpet ja diferentsiaalprivaatsust, võimalusena toodi välja ka plokiahela kasutamine.

Londoni ülikoolist UCL välja kasvanud RegulAItion valiti ka aastal 2020 finantstehnoloogiate häkatoni *Global FinTech Accelerator* võitjaks. Nende pakutud lahendus demonstreeris, kuidas AIR Platform aitaks Aasia finantsasutustel võidelda COVID-19 ja kliimamuutustega seotud probleemidega[79].

### 5.7.4 Sünteetilised andmed ja diferentsiaalprivaatsus statistikas

Ühendkuningriigi statistikaamet ONS on sünteetiliste andmete genereerimist uurinud alates aastast 2018, keskendudes andme- ja masinõppekonveieritele. Alates sellest ajast viiakse läbi projekte, mille eesmärk on võimaldada teaduritel kasutada sünteetilisi andmeid enne juurdepääsu saamist pärisandmetele. 2021. aasta rahvaloenduseks valmistudes genereeris ONS sünteetilisi andmeid koormuse tasakaalustamise ja andmete töötlemisel kasutatavate funktsioonide testimiseks. Sealjuures oli teatud juhtudel oluline, et sünteetilised andmete oleksid sarnase jaotusega nagu pärisandmed. Sünteetilisi andmeid kasutati ka Ühendkuningriigi COVID-19 nakatumise uuringu masinõppeprotsesside vigade silumiseks.

Kui rahvaloenduse kasutusjuhtu loetakse edukaks, siis COVID-19 uuringu sünteetiliste andmete ja pärisandmete statistilised omadused ei läinud piisavalt hästi kokku, mistõttu pärisandmetel tekkinud vead ei tulnud sünteetiliste andmete puhul esile. ONS uurib ka diferentsiaalprivaatsuse kasutamist andmete sünteesimisel, selleks testitakse võistluse NIST *Differential Privacy Challenge* raames välja töötatud meetodeid. Seni läbi viidud projektide järeldusena tuvastas ONS, et lisaks sünteetiliste andmete kasulikkuse ja privaatsusnäitajate hindamisele, tuleb hinnata ka sünteesimisprotsessi arusaadavust. Diferentsiaalprivaatsuse kasutamine võimaldaks andmesünteesi huvipooltel väljendada oma riskitaluvust, aga selleks

[79] RegulAItion wins 2020 Global Fintech Accelerator https://regulaition.com/2020/12/10/regulaition-wins-2020-global-fintech-accelerator (viimati külastatud 02.03.2023).

peab andmete sünteesimise protsess olema arusaadav ka neile, kes pole selle teema spetsialistid [8].

## 5.8 Šveits

### 5.8.1 Nullteadmus ja mikservõrgud internetihääletuses

Šveitsi internetivalimiste ajalugu ulatub sajandivahetusse. Vähene valimisaktiivsus ning laialdased kogemused posti teel kaughääletamisega lõid internetihääletamiseks soodsad olud. Internetihääletamise süsteemi hakati Šveitsis välja töötama mitmes kantonis korraga, esimeste katsetusteni jõuti 2003. aasta referendumil Genfi kantonis. Erinevate süsteemide väljaarendust jätkati detsentraliseeritult, peamiselt kasutasid kantonid enda süsteeme referendumitel [196]. Ilmselt üks enimkasutatud süsteeme on sVote. Algselt ettevõtte Scytl poolt välja arendatud, kuid praegu Šveitsi Posti poolt hallatud süsteemi kasutati kantonites 2016. aastast alates.

Aastal 2019 avastati süsteemis turvaauk, mida seejärel parandama hakati [197]. Aastal 2021 avaldati uusversiooni kood ja dokumentatsioon, et vabatahtlikud saaksid süsteemi testida [198]. Uues süsteemis kasutatakse privaatsuskaitse tehnoloogiatest otspunktkrüpteerimist, mikservõrke ja nullteadmustõestusi. Hääle andmise protsessi alguses hääl krüpteeritakse ning saadetakse hääletusserverile. Enne häälte loendamist kasutatakse nende segamiseks mikservõrke, selle eest vastutavad üksteisest sõltumatud kontrollkomponendid. Mikservõrkude ja ka hiljem häälte dekrüpteerimise juures kasutatakse nullteadmustõestusi, kindlustamaks, et nende protsesside käigus hääli ei lisata, kustutata ega muudeta [199].

### 5.8.2 Hajutatud COVID-19 lähikontaktide tuvastamise süsteem DP-3T

Šveitsi tehnoloogiainstituudi EPFL ja ETH Zürichi teadlased aitasid välja töötada DP-3T tehnoloogiat COVID-19 haigete lähikontaktide tuvastamiseks [200]. Šveitsi lähikontaktsete tuvastamise rakendus SwissCOVID oli ka esimene selle tehnoloogia juurutus. Tehnoloogia rakendas privaatsuse kaitsmiseks krüptograafiat. Erinevalt Singapuri rakendusest *TraceTogether*, mis põhineb *BlueTrace* tehnoloogial, on DP-3T täielikult detsentraliseeritud lahendus. See tähendab, et ajutised võtmed, mida kasutaja seade lähikontakti sattuvate seadmetega vahetab, genereeritakse kasutaja nutitelefonis mittekeskses serveris. Seega saavad vaid kasutajad ise kontrollida, kas nad on nakatanuga kokku puutunud. DP-3T leidis laialdast kasutust avatuse, privaatsusgarantiide ja detsentraliseeritud lähenemise tõttu [193].

### 5.8.3 Homomorfne krüptograafia Šveitsi meditsiiniasutuste võrgus

Šveitsi meditsiiniteaduste akadeemia SAMS koordineerib Šveitsi personaaltervise võrku (*Swiss Personalized Health Network*, SPHN)[80]. Esimese privaatust säilitava ja turvalise terviseandmete jagamise süsteemina võeti kasutusele MedCo[81]. Homomorfse krüptograafia, turvalise ühisarvutuse ja diferentsiaalprivaatsuse abil saavad välised osapooled uuringuid läbi viia mitme asutuse andmete peal [201].

Privaatsuskaitse tehnoloogiate MedCo ja TI4Health võimalikku kasutust uuritakse kahe SPHN teadusprojekti juures – *Swiss Personalized Oncology* ja Swiss BioRef.[82] Esimene on vähiuuringute projekt, mille konsortsiumisse kuuluvad viie ülikooli haiglad ja mitmed regionaalsed vähiravikeskused üle kogu riigi. Teise projekti eesmärgiks on luua võrgustik, mis aitab kokku tuua nelja Šveitsi haigla andmed (enam kui 9 miljonit mõõtmistulemust enam kui 250 000 patsiendi kohta) ja arvutada nende pealt kliiniliste mõõtmiste referentsväärtusi.

[80] https://sphn.ch (viimati külastatud 02.03.2023).
[81] https://medco-ch.github.io/ (viimati külastatud 02.03.2023).
[82] https://sphn.ch/network/projects (viimati külastatud 02.03.2023).

# Krüpteeritud andmete ühisanalüüs Šveitsi personaaltervise võrgus

**Lühidalt:** Šveitsi personaaltervise võrk kasutab privaatsuskaitse tehnoloogiaid süsteemis MedCo, et viia läbi uuringuid mitme terviseasutuse andmete peal.

**Teostamise aasta:** Alates 2018

**Riik:** Šveits

**Omanik:** Šveitsi meditsiiniteaduste akadeemia (SAMS).

**Teostaja:** MedCo väljatöötajaks olid ülikoolid EPFL (*École Spéciale de Lausanne*) ja CHUV (*Lausanne University Hospital*), rahastajateks SPHN (*Swiss Personalised Health Network*) ja PHRT (*Personalised Health and Related Technologies*). Tööstuslikku lahendust TI4Health arendab praegu iduettevõte *Tune Insight*, mis kasvas välja EPFL laboritest.

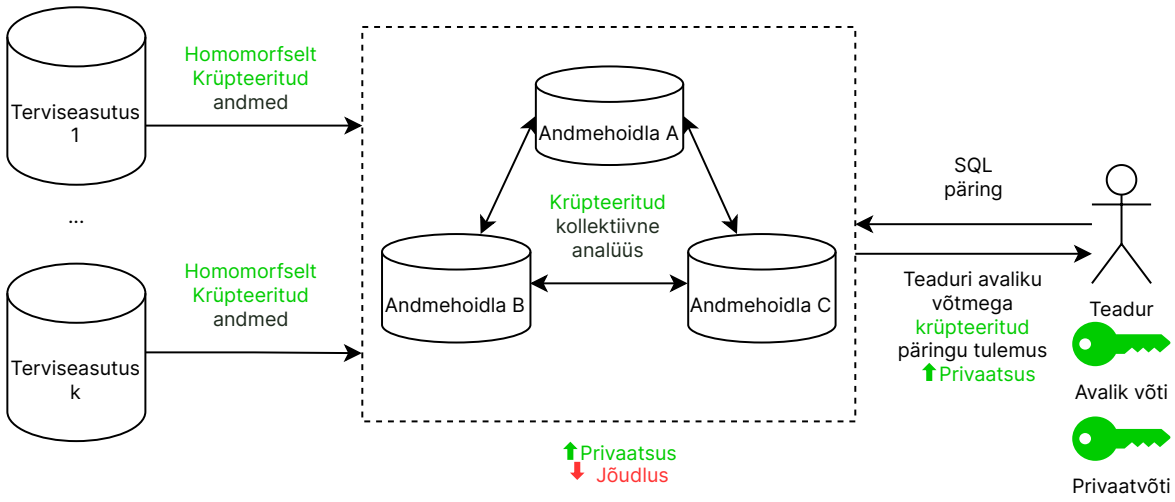
**Süsteemi küpsus:** püsiv juurutus

**Privaatsuskaitse tehnoloogiad:**

1. homomorfne krüptograafia
2. turvaline ühisarvutus
3. diferentsiaalprivaatsus

**Sobivad kasutusjuhtumid:**
privaatne analüütika

**Ülevaatlik mudel:**




**Märkimisväärsed omadused:**

1. MedCo produktsioonivalmis versiooni TI4Health, mida arendab praegu *Tune Insight*, kasutab Šveitsi personaaltervise võrgustik SPHN ning sellel viivad teadlased läbi tundlikele andmetele ligipääsu nõudvaid kliinilisi ja geneetilisi uuringuid.
2. Privaatsuskaitse tehnoloogiate abil teostatakse vähiuuringute projekti ja luuakse nelja Šveitsi haigla andmeid koondavat võrgustikku, mis ühendaks üle 9 miljoni mõõtmistulemuse üle 250 000 patsiendi kohta.

# 6 Privaatsukaitse tehnoloogiate rakendusvõimalused e-riigis

## 6.1 Turvalised andmeruumid e-teenustele

Andmeruumid on mõeldud üksikisikule tema kohta käivate teenuste osutamiseks. Andmeruum on vajalik juhul, kui teenuse osutamiseks vajalikud andmed ei ole ühes andmekogus olemas ning on vaja kokku tuua isiku andmeid mitmest allikast. Andmeruume saavad rakendada nii avaliku kui erasektori teenused.

Andmeruumide kasutusjuhtumeid on mitmeid.

1. Isikule tema sissetulekute ja kulutuste põhjal personaalsete toimetulekutoetuste määramine.
2. Isikule tema sissetulekute põhjal personaalsete trahvide määramine.
3. Isikule tema andmete terviseloo põhjal personaalsete tervisesoovituste andmine.
4. Isikule tema krediidiajaloo põhjal personaalsete krediidipakkumiste tegemine.

Privaatsuskaitse tehnoloogiate abil saab osutada teenust nii, et vajalikke andmeid ei pea tooma kaitsmata kujul ühte kohta kokku, töötlemine toimub kas algusest lõpuni krüpteeritud kujul või riistvaralisi turvameetmeid kasutades. Krüpteeritud kujul koondatud andmete peal saab teha linkimise ning täita päringuid konkreetse isiku kohta.

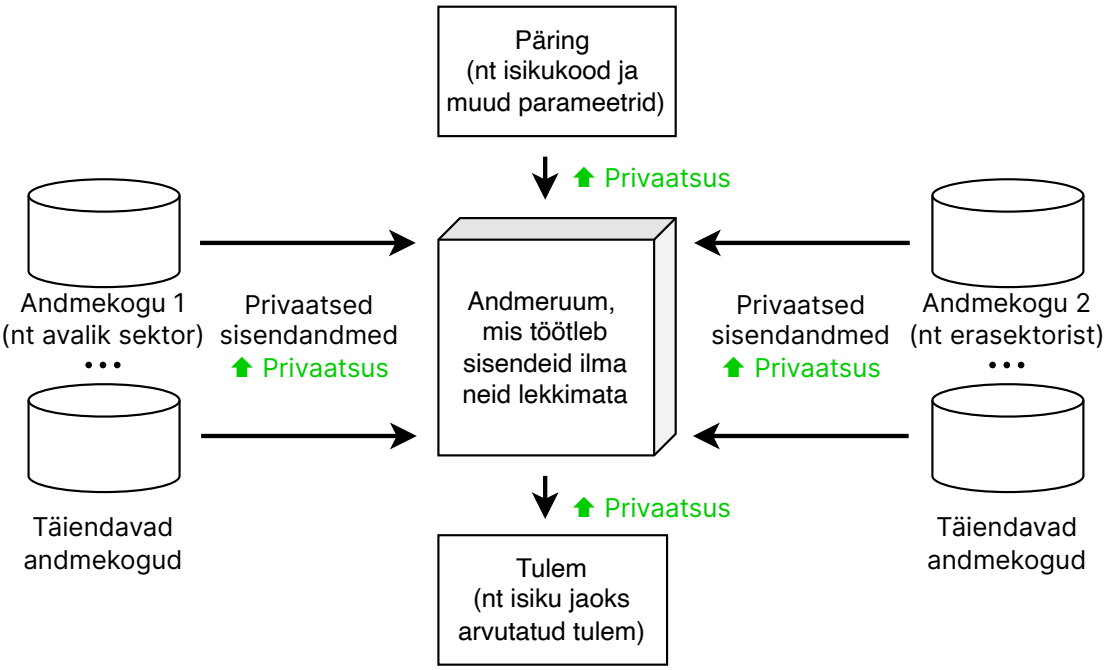
Sellise teenuse eeliseks on tugevam isikuandmete kaitse. Esiteks, vähemad osapooled näevad näevad isiku kohta käivaid andmeid. Teiseks, informatsioon andmeallikate kohta ei ole nähtav andmeruumi teenuseandjale. Andmeruumid võimaldavad ka nähtavuse ja sekkutavuse teenuste teostamist.

# Turvalised andmeruumid teenustele

**Lühidalt:** Mitme (nt avaliku või erasektori) andmebaasi sisu põhjal osutatakse üksikisikule e-teenust.

**Sobivad privaatsuskaitse tehnoloogiad:**
1. usaldatud käivituskeskkonnad (peatükk 4.2.8)
2. turvaline ühisarvutus (peatükk 4.2.10)
3. homomorfne krüptograafia (peatükk 4.2.9)

**Ülevaatlik mudel:**




**Tekkiv lisaväärtus:**
1. Luuakse uusi e-teenuseid, mis kasutavad paremini ära avaliku ja erasektori andmeid ja mida seni oli kõrge riskitaju tõttu keerukam ehitada.
2. Privaatsust kaitsev mõju on eriti suur proaktiivsetele, personaalsetele ja sündmusteenustele, mis on kõrge isikustatuse tasemega.

**Konkreetsed rakendused:**
1. Vajaduspõhiste energia- ja muude toetuste arvutamine
2. Erasektori tervishoiuteenuse osutajate terviseandmete (Apple Health) kaasamine riigi teenustesse
3. Positiivne krediidiregister
4. Digitaalne majutuskaart

**Privaatsuskaitse rakendamise eelised:**
1. Tänu osalisele tsentraliseerimisele on kergem kaasata andmeid osapooltelt, kellel on nt ebamugav juurutada X-tee turvaserverit.
2. Hoolimata tsentraliseerimisest ei teki ühte keskset osapoolt, kes suudaks kõiki andmeid avaldada.

**Näidisrakendused:**
1. Kasutajate kontaktide ühisosa leidmine Signali vestlussüsteemis usaldatavate käivituskeskkondade abil
2. Nõrkade paroolide tuvastamine Microsoft Edge veebilehitsejas homomorfse krüptograafia abil

## 6.2 Privaatne andmete linkimis- ja analüüsiteenus

Turvalise arvutamise abil on võimalik luua ka riiklik andmete ühendamisteenus, mis võimaldaks andmete väljastuste tegemisel neid nt isikukoodide järgi ühendada ning seejärel andmeid taotlenud teadlastele või asutustele üle anda. Sama teenuse edasiarenduses saaks teha ka analüütilist töötlemist, masinõppemudelite treenimist või väärtustamist.

Kui andmeruumid keskenduvad üksikisikule teenuse osutamisele (vt peatükk 6.1), siis andmete linkimise ja analüüsi puhul koondatakse andmeid mitmete isikute kohta. Sellisel juhul pole väljundiks mitte arvutustulemus üksikisiku kohta, vaid statistiline tulem, masinõppemudel või aruanne.

Sellisel moel saab lahendada mitu kasutusjuhtu.

1. Statistiliselt analüüsida andmeruumides osutatud teenuseid ja tehinguid.
2. Koostada aruandeid, statistilisi või masinõppemudeleid mitmest allikast pärit andmebaasidel.
3. Tuvastada pettuseid ja ennetada kuritegusid.

Andmete analüüs toetab ka uute teenuste arendust. Masinõppe rakendused, kratid ning otsusetoe süsteemid, mis töötlevad mitme osapoole isikustatud andmeid, vajavad analüütilist eeltööd nende samade andmete põhjal. Privaatsuskaitse tehnoloogiatega andmete linkimis- ja analüüsiteenus aitab sellist tööd teha turvalisemalt kui tavapärased lahendused.

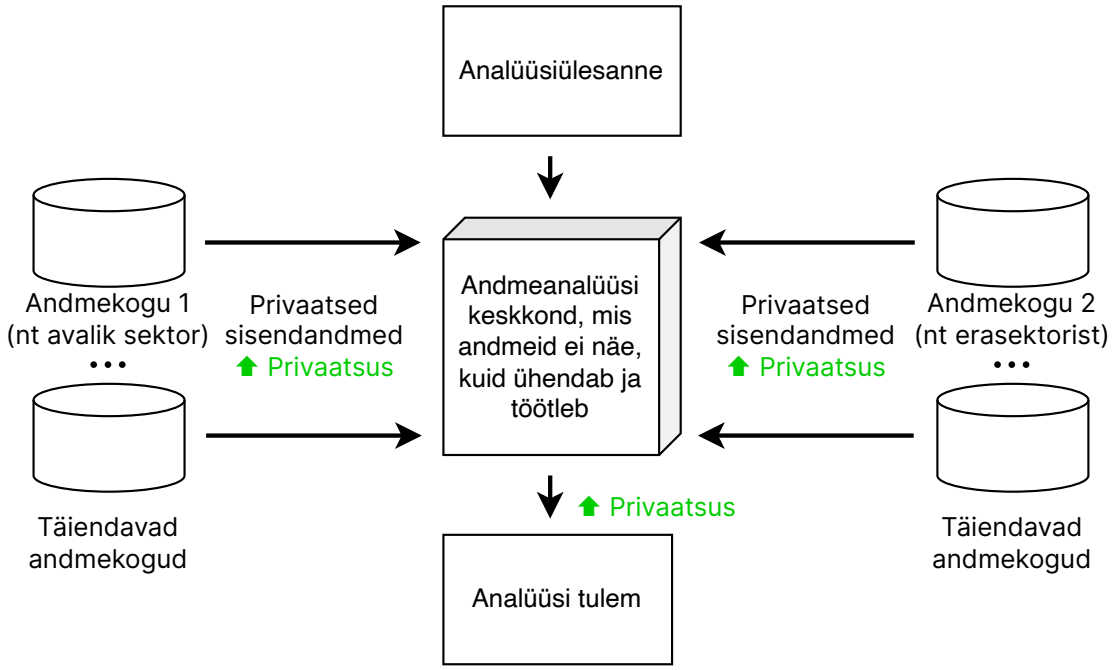
# Privaatne andmete linkimis- ja analüüsiteenus

**Lühidalt:** Mitmest allikast kogutakse andmeid, neid lingitakse ja analüüsitakse privaatsuskaitse tehnoloogiatega, et luua uusi mudeleid ja teadmisi.

**Sobivad privaatsuskaitse tehnoloogiad:**

1. turvaline ühisarvutus (peatükk 4.2.10)
2. usaldatud käivituskeskkonnad (peatükk 4.2.8
3. liitõpe (ei võimalda andmete linkimist tunnuste järgi, ptk 4.2.6)
4. homomorfne krüptograafia (peatükk 4.2.9)
5. täiendavad väljundprivaatsuse tehnikad vastavalt vajadusele

**Ülevaatlik mudel:**




**Tekkiv lisaväärtus:**

1. Luuakse uued analüütikateenused, mis toovad kokku avaliku ja vajadusel ka erasektori andmed

**Konkreetsed rakendused:**

1. Rahapesu või maksupettuste vastase võitluse analüütika
2. Positiivse krediidiregistri analüütika
3. Digitaalsete majutuskaartide analüütika
4. Mobiilside asukohaandmete põhine statistika

**Privaatsuskaitse rakendamise eelised:**

1. Riigi andmete linkimine ja näiteks ka anonüümimine või diferentsiaalprivaatsuse rakendamine on võimalik ära teha enne andmestiku väljastamist.
2. Erasektori andmete kaasamisel on võimalik tõestada, et riik ei saa teha nende andmetega kokkulepitust rohkemat.

**Näidisrakendused:**

1. Bostoni palgalõhe uuring turvalise ühisarvutusega
2. IKT tudengite õpikäitumise uuring maksu- ja haridusandmete peal turvalise ühisarvutusega
3. Suurbritannia piloot toetuspettuste uurimiseks turvalise ühisarvutusega

## 6.3 Avaandmete teenused

Avaandmete teenused saavad väljastada teistele teenustele vajalikke lähteandmeid piiratult, vastavalt päringutele. See võimaldab luua teenuste hierarhiaid, mis toetuvad riigi vastavatele teenustele. Riigil on mitmeid andmeid, mida saavad ära kasutada uute teenuste arendajad. Mitmed neist andmebaasidest ei ole üldse isikutepõhised ning seega ka privaatsuskaitse tehnoloogiaid ei vaja. Küll aga võib privaatsuskaitse tehnoloogiatega olla võimalik avaandmetesse tuua senisest rohkem andmeid. Selle eeldus on, et nende andmete taasisikustamine muudetakse piisavalt keeruliseks.

Isikustatud andmetest tuletatud avaandmete abil saaks ehitada mitmeid uusi teenuseid.

1. Rahvastikustatistikale tuginevad regionaalsed teenused hariduse, tervishoiu, rahanduse, transpordi, energeetika vms valdkondades.
2. Andmeteenused, mis esitlevad inimestele andmeid arusaadavamal ja kasulikumal kujul.

Loodav lisaväärtus tekiks nii ettevõtete kui ka inimeste seas – teadlikkus riigi andmetest ja nende efektiivsem tarbimine tõstaks andmepõhise otsustamise levikut ja juhtimise kvaliteeti.

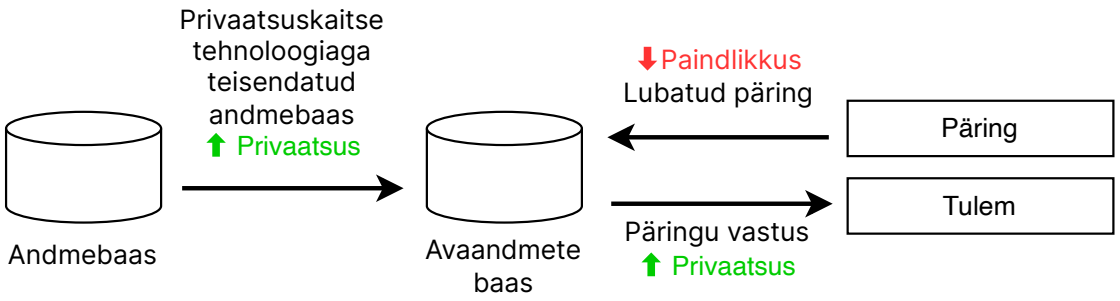
## Avaandmetepõhised e-teenused

**Lühidalt:** Riigi käsutuses oleva andmebaasi vastu lubatakse kasutada teiste teenuste ehitamisel ühe andmeallikana.

**Sobivad privaatsuskaitse tehnoloogiad:**

1. piirangutega päringuliidesed
2. diferentsiaalprivaatsus
3. sünteetilised andmed
4. anonüümimine
5. pseudonüümimine (ei ole isikustatud andmete puhul sobiv)
6. täiendavad privaatsuskaitse tehnoloogiad avaldamise eelseks turvaliseks linkimiseks

**Ülevaatlik mudel:**




**Tekkiv lisaväärtus:**

1. Uued avaandmete teenused seni jagamata peal loovad võimaluse ehitada seni võimatuks (liialt riskantseks) peetud teenuseid.

**Konkreetsed rakendused:**

1. Uued avaandmete teenused näiteks hariduse, tervishoiu, rahanduse, transpordi ja energeetika valdkondades

**Privaatsuskaitse rakendamise eelised:**

1. Privaatsuskaitse tehnoloogiatega kahandatakse riske, et avaandmete teenuse peale ehitatakse mõni ebaeetiline või reputatsiooni rikkuv ärimudel.
2. Privaatsuskaitse tehnoloogiatega kahandatakse avaandmete taasisikustamise riski.

**Näidisrakendused:**

1. Statistikaameti statistika andmebaas
2. Eesti avaandmete portaal
3. Terviseameti COVID-19 anonüümitud avaandmed

## 6.4 Andmebaasi avaldamine avaandmetena

Infosüsteemide, teenuste ja krattide arendamiseks ning testimiseks on vaja andmeid, mis oleksid võimalikult reaalsed, kuid mis poleks isikustatud. Seda põhjusel, et vastavate andmekogude kasutuseesmärkidesse pole testimist tihti ette nähtud. Samuti on teadlastel ja haridusasutustel vaja andmeid, mis kirjeldaks uuritavat valdkonda, kuid ei oleks isikustatud. Sellisel juhul on otstarbekas avaldada terve andmebaas, mis on enne sobivalt töödeldud.

Tegemist on võimsa tööriistaga, millel on palju rakendusi.

1. Andmete kasutamine teadustöös ja poliitikauuringutes.
2. Andmete kasutamine õppetöös, andmeteadlaste õpetamine e-riigi lahendusi ehitama.
3. Infosüsteemide ja analüütikasüsteemide testimine.

Suurim avaandmete kasutamisega seotud oht on seotud vähese kontrolliga nende kasutamise üle. Avaandmete saaja võib taasidentifitseerida isikuid, kasutades selleks vabalt valitud täiendavaid andmebaase, mis talle kättesaadavad on. Seega peab avaandmetega seotud riskianalüüs olema põhjalikum kui kontrollitud keskkondades (nt analüütiku töökohtadel või piiratud päringuliideste kasutamisel).

Teatud juhtudel võib olla põhjendatud andmebaasi üleandmine vaid piiratud hulgale kasutajatele. Siis ei ole küll enam tegemist avaandmetega termini algses tähenduses.

Oluline on ka vahet teha rakendustel, kus on kriitiline seos algandmestiku andmetega (nt seal olevate isikute omadustega) ja kus sellist seost ei ole. Näiteks sotsiaalvaldkonna või tervisestatistikat tehes on avaldatud andmete täpsus oluline. Infosüsteemide testimisel on tähtis, et testimiseks avaldatud andmestik oleks mitmekülgne ja testimisel kasulike eranditega.

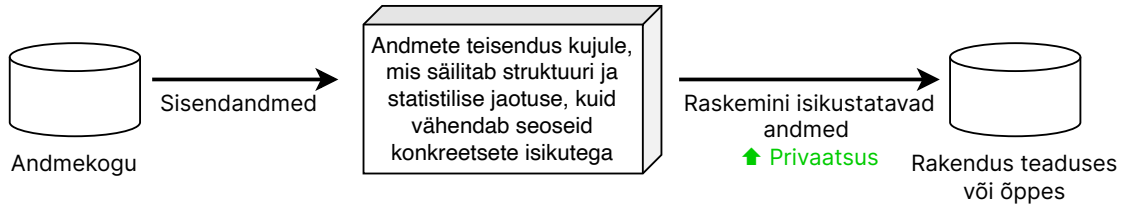
# Andmebaasi avaldamine testimiseks, teadus- või õppetööks

**Lühidalt:** Riigi andmebaas töödeldakse ja antakse kasutajale täies mahus, et seda kasutada õppetöös, teaduses või testimises. Andmebaasi võib teha täiesti avalikuks või anda piiratud hulgale kasutajatele.

**Sobivad privaatsuskaitse tehnoloogiad:**

1. sünteetilised andmed
2. diferentsiaalprivaatsus
3. anonüümimine
4. analüütiku töökohad (sobilik vaid teadus- või õppetööks)
5. pseudonüümimine (ei ole soovitatav)
6. täiendavad privaatsuskaitse tehnoloogiad avaldamise eelseks turvaliseks linkimiseks

**Ülevaatlik mudel:**




**Tekkiv lisaväärtus:**

1. Kvaliteetsete testandmetega kiireneb uute teenuste ja süsteemide loomine ning paraneb nende kvaliteet.
2. Kasvab organisatsioonide arv, kes suudavad luua uusi e-teenuseid.

**Konkreetsed rakendused:**

1. Statistiliste andmete avaldamine
2. Riigi tervishoiu, finants-, ja energiatarbe andmete avaldamine õppe- ja teadustööks

**Privaatsuskaitse rakendamise eelised:**

1. Sõltuvalt valitud tehnoloogiast on avaldatud andmed täiesti juhuslikud, kuid siiski sarnaste statistiliste omadustega.
2. Tugevaid privaatsuskaitse tehnoloogiaid kasutades vähendatakse ohtu asutuse mainele

**Näidisrakendused:**

1. Ameerika Ühendriikide statistikaameti rahvaloenduse andmete avaldamine diferentsiaalprivaatsel kujul
2. LEOSS – Euroopa COVID-19 patsientide andmete jagamine anonüümitult

## 6.5 Sünteetiline digitaalne kaksik riigi andmetest ja teenustest

E-riigi teenuseid tuleb arendada ja hooldada ning selle kvaliteedi tagamiseks on vajalik süsteemide testimine. Testimine tähendab siin nii jõudluse, korrektsuse kui ka kasutatavuse testimist. Selleks on vaja süsteemidesse laadida andmed. Andmekaitse ei luba süsteeme testida isikuandmetega. Päris juhuslike andmetega testimine ei aita aga tagada süsteemi kvaliteeti, sest ei leita üles äärejuhtumeid ja vigu.

Tugevalt teenustele üles ehitatud e-riigi puhul ei piisa testimisel vaid sünteesitud andmebaasidest, vaid vaja on ka teenuseid, mis neid kasutaks. Seega on vaja sünteetilist digitaalset kaksikut peamistest e-teenustest, mille sees oleks sünteesitud andmed.

Sellisel sünteetilisel digitaalsel kaksikul oleks üks peamine kasutus – e-riigi infosüsteemide ja teenuste testimine arenduse käigus. Samas aitaks testandmete probleemi lahendamine märkimisväärselt kaasa süsteemide kvaliteedi ning samal ajal ka andmekaitse ja privaatsuse tagamisele.

Sünteetilise teisiku loomisel tuleb muidugi arvestada, et kvaliteetsete sünteetiliste andmete loomine eeldab lähteandmete analüüsi ja sünteesi mudelite loomist. See tähendab, et lähteandmeid tuleb selleks töödelda. Teadusprojektid on näidanud, et sünteetilisi andmeid saab genereerida ka teiste privaatsuskaitse tehnoloogiatega nagu näiteks usaldatavad käivituskeskkonnad [83].

[83] Andmekaitsenõuetele vastav testandmebaaside süntees turvalise arvutamise tehnoloogiaga. ETAG arendusgrant. https://www.etis.ee/Portal/Projects/Display/6d6fac18-1511-477b-b135-1121d959c903 (viimati külastatud 03.03.2023).

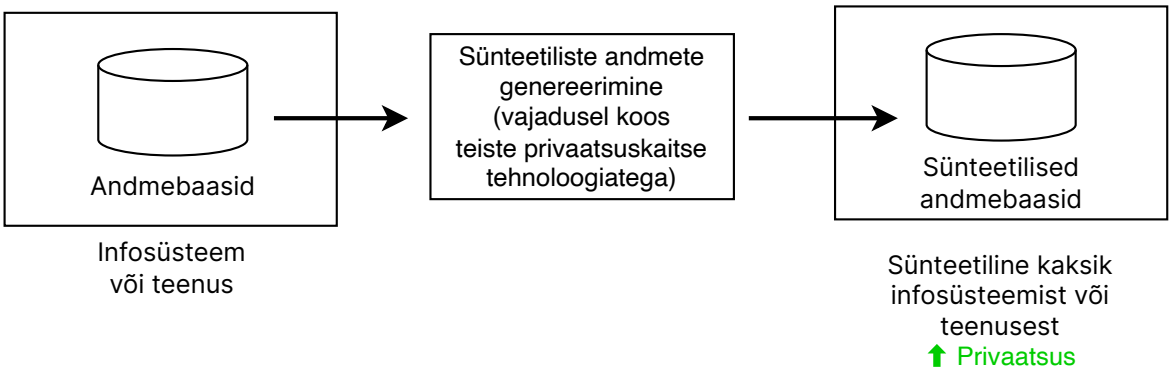
# Sünteetiline digitaalne kaksik riigi andmetest ja teenustest

**Lühidalt:** Riigi teenustest ja nendega seotud andmestikest koostatakse privaatsuskaitse tehnoloogiate abil koopiad, mis ei ole seotud isikutega ning sobivad seega süsteemide testimiseks.

**Sobivad privaatsuskaitse tehnoloogiad:**

1. sünteetiliste andmete genereerimine
2. turvalise arvutamise tehnoloogiad sünteesieelseks lähteandmete turvaliseks linkimiseks

**Ülevaatlik mudel:**




**Tekkiv lisaväärtus:**

1. Kvaliteetsete testandmetega kiireneb uute teenuste ja süsteemide loomine ning paraneb nende kvaliteet.
2. Kasvab organisatsioonide arv, kes suudavad luua uusi e-teenuseid.

**Konkreetsed rakendused:**

1. Sünteetilised digitaalsed kaksikud tervishoiu-, finants- või muudele andmekogudele

**Privaatsuskaitse rakendamise eelised:**

1. Sõltuvalt valitud tehnoloogiast on avaldatud andmed täiesti juhuslikud, kuid siiski sarnaste statistiliste omadustega.
2. Privaatsuskaitse tehnoloogiate abil saab sünteesi käigus lähteandmeid kaitsta.

**Näidisrakendused:**

1. EU-SILC – Euroopa Liidu rahvastikuandmete sünteesimine

## 6.6 Privaatne sündmuste logimine ja logide analüüs

Digitaalsete teenuste kasutamisel on oluline koht logimisel. Logide põhjal saab tuvastada süsteemis esinevaid vigu, anomaaliaid ja väärkasutust. Samal ajal võib süsteemilogi olla väga tundlik andmebaas. Mõelgem näiteks digitaalse identiteedi kasutuse logidele. Kui kasutaja autendib, allkirjastab digitaalset dokumenti või esitab mõnda tõendit, tehakse päring vastava digitaalse identiteedi kehtivuse kontrolli teenusele.

Sellise teenuse pidaja suudab võrguühenduste infot talletades tuletada, milliseid teenuseid isik kasutab ning kui tihti. Teenuse iseloomust (nt mõni tervisenõustamise süsteem, mõni konkreetne veebiteenus) sõltuvalt võib selline logi sisaldada väga tundlikke isikustatavaid andmeid. Euroopa Liidu digitaalse identiteedi kavandites (eIDAS 2.0) on selliste logide pidamisele seatud piirangud. Samas on logid vajalikud pettuste otsinguks ja ennetamiseks.

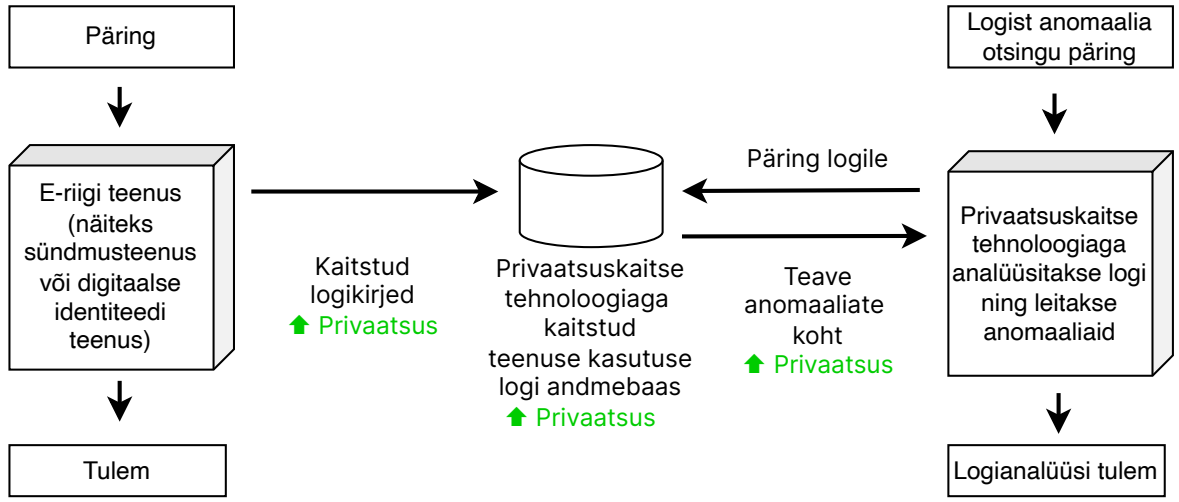
Privaatsuskaitse tehnoloogiate (nt turvaline ühisarvutus, homomorfne krüptograafia ja usaldatavad käivituskeskkonnad) abil on võimalik logisid koguda ning logidest pettuseid ja anomaaliaid leida nii, et teenusepakkujal ei ole logisid lahtisel kujul ning ta ei saa teenuse kasutajat logi kaudu profileerida. Nii on võimalik säilitada logimise funktsionaalsus ja vältida riski, et isiku logide kaudu on võimalik jälgida tema tegevusi.

# Privaatne sündmuste logimine ja logide analüüs

**Lühidalt:** Teenuse kasutuse logi pidamiseks, analüüsiks ja anomaaliate või pettuste otsimiseks kasutatakse privaatsuskaitse tehnoloogiaid. Nii välditakse teenuse kasutajate liigset profileerimist.

**Sobivad privaatsuskaitse tehnoloogiad:**
1. usaldatud käivituskeskkonnad (peatükk 4.2.8)
2. turvaline ühisarvutus (peatükk 4.2.10)
3. homomorfne krüptograafia (peatükk 4.2.9)

**Ülevaatlik mudel:**




**Tekkiv lisaväärtus:**
1. E-riigi teenuste kvaliteet püsib väga hea ilma, et nende kasutajate kohta koguneks liialt põhjalikke profiile.

**Konkreetsed rakendused:**
1. Digitaalse identiteedi kehtivuskinnituste logide kaitse ning anomaaliate tuvastamine.
2. Sündmusteenuste logide kaitse ning anomaaliate tuvastamine.

**Privaatsuskaitse rakendamise eelised:**
1. Uutele teenustele luuakse logid, mis aitavad parandada vigu, leida anomaaliaid ning tuvastada pettuseid.
2. Teenuse logi ei ole taaskasutatav kasutaja profileerimiseks ega sisalda liigseid isikustatud andmeid.

**Näidisrakendused:**
1. Olemasolevaid teostuseid ei ole teada.

## 6.7 Tunnuste ja/või omaduste tõestamine

Euroopa digitaalse identiteedi algatus näeb ette digitaalsete kukrute kasutuselevõttu identiteedi ja tõendite kandjana. Täna levinud tehnoloogiate puhul tähendaks see, et isik peab oma digitaalses kukrus kaasa kandma kõiki tõestatavaid andmeid ning neid ka vajadusel tõenduseks esitama.

Tuleviku kukrud ja isikut tõendavad dokumendid võivad lubada isikul nullteadmustõestuste või nende alamosa, vahemiktõestuste, abil tõestada teenust hankides (näiteks ostu tehes), et ta on vanem kui teatav vanus. Samuti saab rühmaallkirjade abil tõestada, et isik kuulub teatud gruppi või tal on kukrus teatav andmeelement.

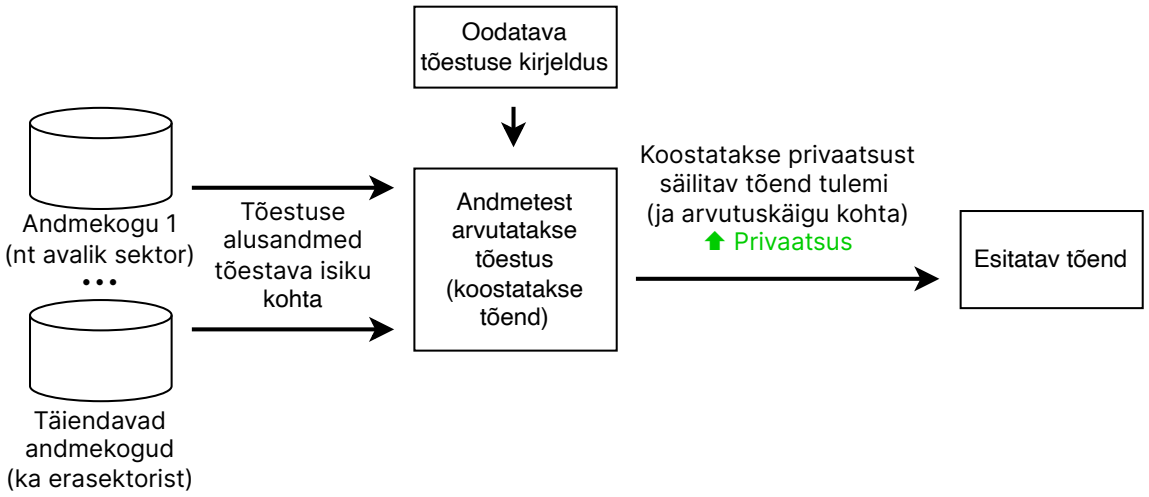
Keerukamate tõendite esitamisel saame kasutada nullteadmustõestuseid, et luua tõendeid, mis ei pane kaasa kogu andmestikku, vaid esitavad kas krüptograafilise tõestuse või otsuse. Näiteks saame teha tõendi, et (a) isiku terviselugu vastab teatud ametiga seotud tervisestandardile, (b) tema pangaarvel on vaid teatud riikide või asutustega tehtud tehingud või (c) tema elektriauto on läbinud teatud vahemaa ning sellest 80% konkreetses riigis. Tõendi kontrollija saab teada, milline on olnud arvutus- ja otsustusprotsess ning kas andmed vastavad nõuetele, kuid ta ei näe lähteandmeid (terviselugu, pangakonto väljavõte, liikuvusandmed).

# Tunnuste ja/või omaduste tõestamine

**Lühidalt:** Tuleviku kukrud ja digitaalne identiteet võib lubada isikul nullteadmustõestuste või vahemiktõestuste abil tõestada poes või teenust hankides, et ta on vanem kui teatav vanus ning saab rühmaallkirjade abil tõestada, et kuulub teatud gruppi või tal on kukrus teatav andmeelement. Tõendite esitamisel saame kasutada nullteadmustõestuseid, et teha tõendeid, mis ei pane kaasa kogu andmestikku, vaid esitab vaid krüptograafilise tõestuse või otsuse, mida tõendi saaja saab kontrollida ja usaldada.

**Sobivad privaatsuskaitse tehnoloogiad:**
1. nullteadmustõestused
2. pimesignatuurid
3. rühma- ja ringisignatuurid

**Ülevaatlik mudel:**




**Tekkiv lisaväärtus:**
1. Uues kukrupõhises digitaalse identiteedi paradigmas on võimalik tagada esitatavate tõendite usaldusväärsus ja vähendada kontrollija vaeva.
2. Luua saab ka riigipiiride üleseid tõendussüsteeme.

**Konkreetsed rakendused:**
1. Vanuse tõestamine
2. Elektriautode läbisõidu ja energiatarbe tõestamine ilma kogu sõiduteekonda esitamata.
3. Tervisestandarditele vastavuse tõestamine ilma terviseandmeid esitamata.
4. Isiku sissetulekute mahu või allikate liigi tõestamine ilma konto väljavõtet avaldamata.

**Privaatsuskaitse rakendamise eelised:**
1. Kuna tõestuse paneb kokku isik omaenda seadmetega, saab ta ka ise vahetult näha enda kohta andmeid (tugevam läbipaistvus).
2. Elementaarsete teenuste osutamise käigus ei pea masinliideste kaudu üle andma terveid andmestikke, piisab vaid tõestusest, et need vastavad soovitud vahemikele või omadustele.

**Näidisrakendused:**
1. Internetihääletamise süsteemid
2. Zcash, Monero, Bytecoin jt krüptovarad

# Bibliograafia

[1] Marit Hansen, Meiko Jensen ja Martin Rost. „Protection Goals for Privacy Engineering“. Teoses: *2015 IEEE Security and Privacy Workshops*, lk. 159–166. DOI: 10.1109/SPW.2015.13.

[2] Dan Bogdanov ja Triin Siil. „Infotehnoloogilised võimalused põhiõiguste kaitsel“. *Juridica* 6 (2020), lk. 474–481. URL: https://juridica.ee/article.php?uri=2020_6_infotehnoloogilised_v_imalused_p_hi_iguste_kaitsel.

[3] *Euroopa Parlamendi ja nõukogu määrus (EL) 2016/679, 27. aprill 2016, füüsiliste isikute kaitse kohta isikuandmete töötlemisel ja selliste andmete vaba liikumise ning direktiivi 95/46/EÜ kehtetuks tunnistamise kohta (isikuandmete kaitse üldmäärus)*. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[4] Johanna Vallistu, Tea Danilov ja Uku Varblane. *Andmeühiskonna tulevik. Stsenaariumid aastani 2035*. Tehniline raport. Arenguseire Keskus, 2022. URL: https://arenguseire.ee/wp-content/uploads/2022/12/2022_andmeyhiskonna-tulevik_raport.pdf.

[5] *Eesti digiühiskond 2030. Valdkonna arengukava*. Tehniline raport. Eesti Vabariigi Majandus- ja Kommunikatsiooniministeerium, 2021. URL: https://mkm.ee/digiriik-ja-uhenduvus/digiuhiskonna-arengukava-2030.

[6] The Royal Society. *From privacy to partnership: The role of privacy enhancing technologies in data governance and collaborative analysis*. 2023. URL: https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/From-Privacy-to-Partnership.pdf?la=en-GB&hash=4769FEB5C984089FAB52FE7E22F379D6.

[7] Centre for Data Ethics ja Innovation's (CDEI). *Privacy Enhanching Technologies Adoption Guide*. 2021. URL: https://cdeiuk.github.io/pets-adoption-guide.

[8] United Nations Committee of Experts on Big Data ja Data Science for Official Statistics. *United Nations Guide on Privacy-Enhancing Technologies for Official Statistics*. United Nations, 2023. URL: https://unstats.un.org/bigdata/task-teams/privacy/guide/2023_UN%20PET%20Guide.pdf.

[9] Euroopa Komisjon. *KOMISJONI TEATIS EUROOPA PARLAMENDILE, NÕUKOGULE, EUROOPA MAJANDUS- JA SOTSIAALKOMITEELE NING REGIOONIDE KOMITEELE Euroopa andmestrateegia*. URL: https://eur-lex.europa.eu/legal-content/ET/TXT/?uri=CELEX:52020DC0066.

[10] Euroopa Parlament ja Euroopa Liidu Nõukogu. *Euroopa Parlamendi ja nõukogu määrus (EL) 2022/868, 30. mai 2022, Euroopa andmehalduse kohta ning millega muudetakse määrust (EL) 2018/1724 (andmehalduse määrus)*. URL: https://data.europa.eu/eli/reg/2022/868/oj.

[11] Euroopa Parlament ja Euroopa Liidu Nõukogu. *Euroopa Parlamendi ja nõukogu määrus (EL) 2022/1925, 14. september 2022, mis käsitleb konkurentsile avatud ja õiglaseid turge digisektoris ning millega muudetakse direktiive (EL) 2019/1937 ja (EL) 2020/1828 (digiturgude määrus)*. URL: https://data.europa.eu/eli/reg/2022/1925/oj.

[12] Euroopa Andmekaitsenõukogu. *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default Version 2.0 adopted on 20 October 2020*. URL: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf.

[13] Information Commissioner's Office (ICO). *Chapter 5: Privacy-enhancing technologies (PETs). Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance*. 2020. URL: https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf.

[14] *Information technology — Security techniques — Privacy engineering for system life cycle processes*. Standard ISO/IEC TR 27550:2019. 2019. URL: https://www.iso.org/standard/72024.html.

[15] *Systems and software engineering — Software life cycle processes*. Standard ISO/IEC/IEEE 12207:2017. 2017. URL: https://www.iso.org/standard/63712.html.

[16] *Systems and software engineering — Life cycle processes — Requirements engineering*. Standard ISO/IEC/IEEE 29148:2018. 2018. URL: https://www.iso.org/standard/72089.html.

[17] *Systems and software engineering — System life cycle processes*. Standard ISO/IEC/IEEE 15288:2015. ISO/IEC/IEEE, 2015. URL: https://www.iso.org/standard/72089.html.

[18] Andmekaitse Inspektsioon. *Isikuandmete töötleja üldjuhend (2019)*. URL: https://www.aki.ee/sites/default/files/dokumendid/isikuandmete_tootleja_uldjuhend.pdf.

[19] Artikli 29 alusel asutatud andmekaitse töörühm. *Suunised, mis käsitlevad andmekaitsealast mõjuhinnangut ja selle kindlaksmääramist, kas isikuandmete töötlemise tulemusena „tekib tõenäoliselt suur oht" vastavalt määrusele (EL) 2016/679*. 2017.

[20] Pille Pullonen, Raimundas Matulevičius ja Dan Bogdanov. „PE-BPMN: Privacy-Enhanced Business Process Model and Notation". Teoses: *Business Process Management*. Springer International Publishing, 2017, lk. 40–56. DOI: 10.1007/978-3-319-65000-5_3.

[21] Marlon Dumas *et al*. „Multi-level privacy analysis of business processes: the Pleak toolset". Teoses: *International Journal on Software Tools for Technology Transfer*. Köide 24. 2. 2022, lk. 183–203. DOI: 10.1007/s10009-021-00636-w.

[22] *Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management — Requirements and guidelines*. Standard ISO/IEC 27701:2019. 2019. URL: https://www.iso.org/standard/71670.html.

[23] *Eesti infoturbestandard*. URL: https://eits.ria.ee/.

[24] OASIS. *Privacy Management Reference Model and Methodology (PMRM) Version 1.0 Committee Specification 02*. 2016. URL: https://docs.oasis-open.org/pmrm/PMRM/v1.0/cs02/PMRM-v1.0-cs02.pdf.

[25] Rick Kazman, Mark Klein ja Paul Clements. *ATAM: Method for architecture evaluation*. Tehniline raport. Carnegie-Mellon Univ Pittsburgh PA Software Engineering Inst, 2000. URL: https://apps.dtic.mil/sti/citations/ADA382629.

[26] Kim Wuyts ja Wouter Joosen. *LINDDUN privacy threat modeling: a tutorial*. Tehniline raport. 2015. URL: https://www.cs.kuleuven.be/publicaties/rapporten/cw/CW685.abs.html.

[27] *Information technology — Security techniques — Guidelines for privacy impact assessment*. Standard ISO/IEC 29134:2017. 2017. URL: https://www.iso.org/standard/62289.html.

[28] Joint Task Force. „Assessing security and privacy controls in information systems and organizations“. *NIST Special Publication* (2022). DOI: 10.6028/NIST.SP.800-53Ar5.

[29] *Information security, cybersecurity and privacy protection — Information security management systems — Requirements*. Standard ISO/IEC 27001:2022. 2022. URL: https://www.iso.org/standard/82875.html.

[30] European Union Agency for Cybersecurity *et al. Pseudonymisation techniques and best practices: recommendations on shaping technology according to data protection and privacy provisions*. 2019. DOI: 10.2824/247711.

[31] European Union Agency for Cybersecurity *et al. Data pseudonymisation: advanced techniques and use cases: technical analysis of cybersecurity measures in data protection and privacy*. European Union Agency for Cybersecurity (ENISA), 2022. DOI: 10.2824/860099.

[32] Valentina Ciriani *et al.* „$\kappa$-anonymity“. *Secure data management in decentralized systems* (2007), lk. 323–353. DOI: 10.1007/978-0-387-27696-0_10.

[33] Carolin E. M. Jakob *et al.* „Design and evaluation of a data anonymization pipeline to promote Open Science on COVID-19“. *Sci Data* 7.1 (2020), lk. 435. DOI: 10.1038/s41597-020-00773-y.

[34] Chris Culnane, Benjamin I. P. Rubinstein ja Vanessa Teague. „Stop the Open Data Bus, We Want to Get Off“. *CoRR* (2019). URL: https://arxiv.org/abs/1908.05004.

[35] Luc Rocher, Julien M. Hendrickx ja Yves-Alexandre De Montjoye. „Estimating the success of re-identifications in incomplete datasets using generative models“. *Nature communications* 10.1 (2019), lk. 1–9.

[36] Dmitry Prokhorenkov. „Anonymization Level and Compliance for Differential Privacy: A Systematic Literature Review“. Teoses: *2022 International Wireless Communications and Mobile Computing (IWCMC)*. 2022, lk. 1119–1124. DOI: 10.1109/IWCMC55113.2022.9824899.

[37] Ana-Maria Cretu *et al.* „QuerySnout: Automating the Discovery of Attribute Inference Attacks against Query-Based Systems“. Teoses: CCS '22. Los Angeles, CA, USA: Association for Computing Machinery, 2022, lk. 623–637. DOI: 10.1145/3548606.3560581.

[38] Simson Garfinkel. „Differential Privacy and the 2020 US Census“. *MIT Case Studies in Social and Ethical Responsibilities of Computing* Winter 2022 (). URL: https://mit-serc.pubpub.org/pub/differential-privacy-2020-us-census.

[39] Apple Differential Privacy Team. *Learning with privacy at scale*. 2017. URL: https://machinelearning.apple.com/research/learning-with-privacy-at-scale.

[40] Úlfar Erlingsson, Vasyl Pihur ja Aleksandra Korolova. „RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response“. Teoses: *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*. ACM, lk. 1054–1067. DOI: 10.1145/2660267.2660348.

[41] Cynthia Dwork ja Aaron Roth. „The Algorithmic Foundations of Differential Privacy“. *Found. Trends Theor. Comput. Sci.* 9.3-4 (2014), lk. 211–407. DOI: 10.1561/0400000042.

[42] Cynthia Dwork *et al*. „Calibrating Noise to Sensitivity in Private Data Analysis“. Teoses: *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*. Köide 3876. Lecture Notes in Computer Science. Springer, lk. 265–284. DOI: 10.1007/11681878_14.

[43] Irit Dinur ja Kobbi Nissim. „Revealing information while preserving privacy“. Teoses: *Proceedings of the Twenty-Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 9-12, 2003, San Diego, CA, USA*. ACM, lk. 202–210. DOI: 10.1145/773153.773173.

[44] Jon M. Kleinberg, Christos H. Papadimitriou ja Prabhakar Raghavan. „Auditing Boolean attributes“. *J. Comput. Syst. Sci.* 66.1 (2003), lk. 244–253. DOI: 10.1016/S0022-0000(02)00036-3.

[45] Alisa Pankova ja Peeter Laud. „Interpreting Epsilon of Differential Privacy in Terms of Advantage in Guessing or Approximating Sensitive Attributes“. Teoses: *35th IEEE Computer Security Foundations Symposium, CSF 2022, Haifa, Israel, August 7-10, 2022*. IEEE, 2022, lk. 96–111. DOI: 10.1109/CSF54842.2022.9919656.

[46] Peeter Laud, Alisa Pankova ja Martin Pettai. „A Framework of Metrics for Differential Privacy from Local Sensitivity“. *Proc. Priv. Enhancing Technol.* 2020.2 (2020), lk. 175–208. DOI: 10.2478/popets-2020-0023.

[47] Dan Zhang *et al*. „EKTELO: A Framework for Defining Differentially-Private Computations“. Teoses: *Proceedings of the 2018 International Conference on Management of Data, SIGMOD Conference 2018, Houston, TX, USA, June 10-15, 2018*. ACM, lk. 115–130. DOI: 10.1145/3183713.3196921.

[48] Rachel Cummings ja Deven Desai. „The role of differential privacy in gdpr compliance“. Teoses: *FAT'18: Proceedings of the Conference on Fairness, Accountability, and Transparency*. 2018, lk. 20. URL: https://cpb-us-w2.wpmucdn.com/sites.gatech.edu/dist/c/679/files/2018/09/GDPR_DiffPrivacy.pdf.

[49] Kobbi Nissim *et al*. „Bridging the gap between computer science and legal approaches to privacy“. *Harv. JL & Tech.* 31 (2017), lk. 687. URL: https://privacytools.seas.harvard.edu/files/privacytools/files/02.-article-wood-7.21.pdf.

[50] Andrew Hard *et al*. „Federated Learning for Mobile Keyboard Prediction“. *CoRR* (2018). URL: http://arxiv.org/abs/1811.03604.

[51] Swaroop Ramaswamy *et al*. „Federated Learning for Emoji Prediction in a Mobile Keyboard“. *CoRR* (2019). URL: http://arxiv.org/abs/1906.04329.

[52] Matthias Paulik *et al*. „Federated Evaluation and Tuning for On-Device Personalization: System Design & Applications“. *CoRR* (2021). URL: https://arxiv.org/abs/2102.08503.

[53] Dzmitry Huba *et al*. „Papaya: Practical, Private, and Scalable Federated Learning“. *CoRR* (2021). URL: https://arxiv.org/abs/2111.04877.

[54] Matthew Fredrikson *et al*. „Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing“. Teoses: *23rd USENIX Security Symposium (USENIX Security 14)*. San Diego, CA: USENIX Association, 2014, lk. 17–32. URL: https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/fredrikson_matthew.

[55] Kallista A. Bonawitz *et al*. „Towards Federated Learning at Scale: System Design“. *CoRR* (2019). URL: http://arxiv.org/abs/1902.01046.

[56] Brendan McMahan ja Abhradeep Thakurta. *Federated Learning with Formal Differential Privacy Guarantees*. URL: https://ai.googleblog.com/2022/02/federated-learning-with-formal.html.

[57] Eurostat. *EU Statistics on Income and Living Conditions (EU-SILC)*. URL: https://ec.europa.eu/eurostat/web/microdata/statistics-on-income-and-living-conditions.

[58] César Augusto Fontanillo López ja Abdullah Elbi. *On synthetic data: a brief introduction for data protection law dummies*. 2022. URL: https://europeanlawblog.eu/2022/09/22/on-synthetic-data-a-brief-introduction-for-data-protection-law-dummies.

[59] Jaak Randmets. *An Overview of Vulnerabilities and Mitigations of Intel SGX Applications*. Tehniline raport D-2-116. C, 2021. URL: https://cyber.ee/research/reports.

[60] Muhammad Naveed, Seny Kamara ja Charles V. Wright. „Inference Attacks on Property-Preserving Encrypted Databases". Teoses: *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. CCS '15. Denver, Colorado, USA: Association for Computing Machinery, 2015, lk. 644–655. DOI: 10.1145/2810103.2813651.

[61] Ronald L. Rivest, Adi Shamir ja Leonard Adleman. „A method for obtaining digital signatures and public-key cryptosystems". *Communications of the ACM* 21.2 (1978), lk. 120–126. DOI: 10.1145/359340.359342.

[62] Pascal Paillier. „Public-key cryptosystems based on composite degree residuosity classes". Teoses: *Advances in Cryptology—EUROCRYPT'99: International Conference on the Theory and Application of Cryptographic Techniques Prague, Czech Republic, May 2–6, 1999 Proceedings 18*. Springer, 1999, lk. 223–238. DOI: 10.1007/3-540-48910-X_16.

[63] Craig Gentry. „Fully Homomorphic Encryption Using Ideal Lattices". Teoses: *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*. STOC '09. Bethesda, MD, USA: Association for Computing Machinery, 2009, lk. 169–178. DOI: 10.1145/1536414.1536440.

[64] ISO Central Secretary. *Information security — Secure multiparty computation — Part 1: General*. Standard ISO/IEC DIS 4922-1. Geneva, CH, 2022. URL: https://www.iso.org/standard/80508.html.

[65] Usable ja Efficient Secure Multiparty Computation (UaESMC) project. *Deliverable D1.1: Capability Model*. Tehniline raport. 2012. URL: https://uaesmc.cyber.ee/files/D1.1.pdf.pdf.

[66] David W. Archer *et al*. *From Keys to Databases – Real-World Applications of Secure Multi-Party Computation*. 2018. DOI: 10.1093/comjnl/bxy090.

[67] Riivo Talviste. „Applying secure multi-party computation in practice". *Ph. D. dissertation* (2016). URL: https://hdl.handle.net/10062/50510.

[68] Dan Bogdanov *et al*. „Students and Taxes: a Privacy-Preserving Study Using Secure Computation". *Proc. Priv. Enhancing Technol.* 2016.3 (2016), lk. 117–135. DOI: 10.1515/popets-2016-0019.

[69] Andrew Chi-Chih Yao. „How to generate and exchange secrets". Teoses: *27th Annual Symposium on Foundations of Computer Science (sfcs 1986)*. 1986, lk. 162–167. DOI: 10.1109/SFCS.1986.25.

[70] Dahlia Malkhi *et al*. „Fairplay – Secure Two-Party Computation System“. Teoses: *13th USENIX Security Symposium (USENIX Security 04)*. 2004, lk. 287–302. URL: https://www.usenix.org/legacy/event/sec04/tech/full_papers/malkhi/malkhi_html.

[71] Peter Bogetoft *et al*. „Secure Multiparty Computation Goes Live“. Teoses: *13th International Conference of Financial Cryptography and Data Security. FC'09*. 2009, lk. 325–343. DOI: 10.1007/978-3-642-03549-4_20.

[72] Dan Bogdanov, Riivo Talviste ja Jan Willemson. „Deploying secure multi-party computation for financial data analysis (short paper)“. Teoses: *Proceedings of the 16th International Conference on Financial Cryptography and Data Security. FC'12*. 2012, lk. 57–64. DOI: 10.1007/978-3-642-32946-3_5.

[73] Dan Bogdanov *et al*. „Rmind: A Tool for Cryptographically Secure Statistical Analysis“. *IEEE Transactions on Dependable and Secure Computing* 15.3 (2018), lk. 481–495. DOI: 10.1109/TDSC.2016.2587623.

[74] Marcella Hastings *et al*. „SoK: General Purpose Compilers for Secure Multi-Party Computation“. Teoses: *2019 IEEE Symposium on Security and Privacy, San Francisco, CA, USA, May 19-23, 2019*. IEEE, lk. 1220–1237. DOI: 10.1109/SP.2019.00028.

[75] Gerald Spindler ja Philipp Schmechel. „Personal Data and Encryption in the European General Data Protection Regulation“. *JIPITEC* 7.2 (2016), lk. 163–177. URL: https://heinonline.org/HOL/LandingPage?handle=hein.journals/jipitec7&div=18&id=&page=.

[76] *Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data*. Tehniline raport. European Data Protection Board, 2020. URL: https://edpb.europa.eu/sites/default/files/consultation/edpb_recommendations_202001_supplementarymeasurestransferstools_en.pdf.

[77] David Chaum. „Blind Signature System“. Teoses: *Advances in Cryptology, Proceedings of CRYPTO '83, Santa Barbara, California, USA, August 21-24, 1983*. Plenum Press, New York, 1983, lk. 153. DOI: 10.1007/978-1-4684-4730-9_14.

[78] David Chaum. „Blind Signatures for Untraceable Payments“. Teoses: *Advances in Cryptology: Proceedings of CRYPTO '82, Santa Barbara, California, USA, August 23-25, 1982*. Plenum Press, New York, 1982, lk. 199–203. DOI: 10.1007/978-1-4757-0602-4_18.

[79] David Chaum, Amos Fiat ja Moni Naor. „Untraceable Electronic Cash“. Teoses: *Advances in Cryptology - CRYPTO '88, 8th Annual International Cryptology Conference, Santa Barbara, California, USA, August 21-25, 1988, Proceedings*. Köide 403. Lecture Notes in Computer Science. Springer, 1988, lk. 319–327. DOI: 10.1007/0-387-34799-2_25.

[80] Markus Stadler, Jean-Marc Piveteau ja Jan Camenisch. „Fair Blind Signatures“. Teoses: *Advances in Cryptology - EUROCRYPT '95, International Conference on the Theory and Application of Cryptographic Techniques, Saint-Malo, France, May 21-25, 1995, Proceeding*. Köide 921. Lecture Notes in Computer Science. Springer, 1995, lk. 209–219. DOI: 10.1007/3-540-49264-X_17.

[81] Jan Camenisch ja Anna Lysyanskaya. „An Efficient System for Non-transferable Anonymous Credentials with Optional Anonymity Revocation“. Teoses: *Advances in Cryptology - EUROCRYPT 2001, International Conference on the Theory and Application of Cryptographic Techniques, Innsbruck, Austria, May 6-10, 2001, Proceeding*. Köide 2045. Lecture Notes in Computer Science. Springer, 2001, lk. 93–118. DOI: 10.1007/3-540-44987-6_7.

[82] Alwyn Goh ja W. K. Yip. „A divisible extension of the Brands digital cash protocol: k-term coins implemented via secret sharing“. Teoses: *2000 TENCON Proceedings. Intelligent Systems and Technologies for the New Millennium (Cat. No.00CH37119)*. Köide 3. 2000, lk. 452–457. DOI: 10.1109/TENCON.2000.892308.

[83] Tatsuaki Okamoto ja Kazuo Ohta. „Disposable Zero-Knowledge Authentications and Their Applications to Untraceable Electronic Cash“. Teoses: *Advances in Cryptology - CRYPTO '89, 9th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 1989, Proceedings*. Köide 435. Lecture Notes in Computer Science. Springer, 1989, lk. 481–496. DOI: 10.1007/0-387-34805-0_43.

[84] Foteini Baldimtsi *et al.* „Anonymous Transferable E-Cash“. Teoses: *Public-Key Cryptography - PKC 2015 - 18th IACR International Conference on Practice and Theory in Public-Key Cryptography, Gaithersburg, MD, USA, March 30 - April 1, 2015, Proceedings*. Köide 9020. Lecture Notes in Computer Science. Springer, 2015, lk. 101–124. DOI: 10.1007/978-3-662-46447-2_5.

[85] Paul Schmitt ja Barath Raghavan. „Pretty Good Phone Privacy“. Teoses: *30th USENIX Security Symposium (USENIX Security 2021)*. USENIX Association, 2021, lk. 1737–1754. URL: https://www.usenix.org/conference/usenixsecurity21/presentation/schmitt.

[86] CNET. *Google One VPN: What you need to know about this privacy tool*. URL: https://www.cnet.com/tech/services-and-software/google-one-vpn-what-you-need-to-know-about-this-privacy-tool.

[87] Matthew Green ja Ian Miers. „Bolt: Anonymous Payment Channels for Decentralized Currencies“. Teoses: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017, Dallas, TX, USA, October 30 - November 03, 2017*. ACM, lk. 473–489. DOI: 10.1145/3133956.3134093.

[88] Anna Lysyanskaya. *Security Analysis of RSA-BSSA*. Cryptology ePrint Archive, Paper 2022/895. 2022. URL: https://eprint.iacr.org/2022/895.

[89] Jakob Jonsson ja Burt Kaliski. *Public-Key Cryptography Standards (PKCS) #1: RSA Cryptography Specifications Version 2.1*. RFC 3447. 2003. URL: https://rfc-editor.org/rfc/rfc3447.txt.

[90] David Chaum ja Eugène van Heyst. „Group Signatures“. Teoses: *Advances in Cryptology - EUROCRYPT '91, Workshop on the Theory and Application of of Cryptographic Techniques, Brighton, UK, April 8-11, 1991, Proceedings*. Köide 547. Lecture Notes in Computer Science. Springer, lk. 257–265. DOI: 10.1007/3-540-46416-6_22.

[91] Ronald L. Rivest, Adi Shamir ja Yael Tauman. „How to Leak a Secret“. Teoses: *Advances in Cryptology - ASIACRYPT 2001, 7th International Conference on the Theory and Application of Cryptology and Information Security, Gold Coast, Australia, December 9-13, 2001, Proceedings*. Köide 2248. Lecture Notes in Computer Science. Springer, lk. 552–565. DOI: 10.1007/3-540-45682-1_32.

[92] Fumitaka Hoshino, Tetsutaro Kobayashi ja Koutarou Suzuki. „Anonymizable Signature and Its Construction from Pairings“. Teoses: *Pairing-Based Cryptography - Pairing 2010 - 4th International Conference, Yamanaka Hot Spring, Japan, December 2010. Proceedings*. Köide 6487. Lecture Notes in Computer Science. Springer, lk. 62–77. DOI: 10.1007/978-3-642-17455-1_5.

[93] Moneropedia. *Ring Signature*. URL: https://www.getmonero.org/resources/moneropedia/ringsignatures.html.

[94] Moneropedia. *CLSAG*. URL: https://www.getmonero.org/resources/moneropedia/clsag.html.

[95] Bruhadeshwar Bezawada ja Indrakshi Ray. „Attribute-Based Encryption: Applications and Future Directions“. Teoses: *From Database to Cyber Security - Essays Dedicated to Sushil Jajodia on the Occasion of His 70th Birthday*. Köide 11170. Lecture Notes in Computer Science. Springer, 2018, lk. 353–374. DOI: 10.1007/978-3-030-04834-1_18.

[96] ETSI Technical Committee Cyber Security (CYBER). *Attribute Based Encryption for Attribute Based Access Control*. Standard ETSI TS 103 532. Sophia Antipolis: European Telecommunications Standards Institute (ETSI), 2021. URL: https://www.etsi.org/deliver/etsi_ts/103500_103599/103532/01.02.01_60/ts_103532v010201p.pdf.

[97] ETSI Technical Committee Cyber Security (CYBER). *Application of Attribute Based Encryption (ABE) or PII and personal data protection on IoT devices, WLAN, cloud and mobile services – High level requirements*. Standard ETSI TS 103 458. Sophia Antipolis: European Telecommunications Standards Institute (ETSI), 2018. URL: https://www.etsi.org/deliver/etsi_ts/103400_103499/103458/01.01.01_60/ts_103458v010101p.pdf.

[98] Vipul Goyal *et al*. „Attribute-based encryption for fine-grained access control of encrypted data“. Teoses: *Proceedings of the 13th ACM Conference on Computer and Communications Security, CCS 2006, Alexandria, VA, USA, October 30 - November 3, 2006*. ACM, lk. 89–98. DOI: 10.1145/1180405.1180418.

[99] John Bethencourt, Amit Sahai ja Brent Waters. „Ciphertext-Policy Attribute-Based Encryption“. Teoses: *2007 IEEE Symposium on Security and Privacy (S&P 2007), 20-23 May 2007, Oakland, California, USA*. IEEE Computer Society, lk. 321–334. DOI: 10.1109/SP.2007.11.

[100] Aleksandr Lenin ja Peeter Laud. „Attribute-based encryption for named data networking“. Teoses: *ICN '21: 8th ACM Conference on Information-Centric Networking, Paris, France, September 22 - 24, 2021*. ACM, lk. 118–120. DOI: 10.1145/3460417.3483371.

[101] Amit Sahai ja Brent Waters. „Fuzzy Identity-Based Encryption“. Teoses: *Advances in Cryptology - EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, Proceedings*. Köide 3494. Lecture Notes in Computer Science. Springer, lk. 457–473. DOI: 10.1007/11426639_27.

[102] Adi Shamir. „Identity-Based Cryptosystems and Signature Schemes“. Teoses: *Advances in Cryptology, Proceedings of CRYPTO '84, Santa Barbara, California, USA, August 19-22, 1984, Proceedings*. Köide 196. Lecture Notes in Computer Science. Springer, lk. 47–53. DOI: 10.1007/3-540-39568-7_5.

[103] Dan Boneh ja Matthew K. Franklin. „Identity-Based Encryption from the Weil Pairing“. *SIAM J. Comput.* 32.3 (2003), lk. 586–615. DOI: 10.1137/S0097539701398521.

[104] Allison B. Lewko ja Brent Waters. „Decentralizing Attribute-Based Encryption“. Teoses: *Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings*. Köide 6632. Lecture Notes in Computer Science. Springer, lk. 568–588. DOI: 10.1007/978-3-642-20465-4_31.

[105] Hemanta K. Maji, Manoj Prabhakaran ja Mike Rosulek. „Attribute-Based Signatures“. Teoses: *Topics in Cryptology - CT-RSA 2011 - The Cryptographers' Track at the RSA Conference 2011, San Francisco, CA, USA, February 14-18, 2011. Proceedings*. Köide 6558. Lecture Notes in Computer Science. Springer, lk. 376–392. DOI: 10.1007/978-3-642-19074-2_24.

[106] Ronald Cramer ja Ivan Damgård. „Zero-Knowledge Proofs for Finite Field Arithmetic; or: Can Zero-Knowledge be for Free?“ Teoses: *Advances in Cryptology - CRYPTO '98, 18th Annual International Cryptology Conference, Santa Barbara, California, USA, August 23-27, 1998, Proceedings*. Köide 1462. Lecture Notes in Computer Science. Springer, lk. 424–441. DOI: 10.1007/BFb0055745.

[107] Kang Yang *et al.* „QuickSilver: Efficient and Affordable Zero-Knowledge Proofs for Circuits and Polynomials over Any Field“. Teoses: *CCS '21: 2021 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, Republic of Korea, November 15 - 19, 2021*. ACM, lk. 2986–3001. DOI: 10.1145/3460120.3484556.

[108] Carsten Baum *et al.* „Mac'n'Cheese: Zero-Knowledge Proofs for Boolean and Arithmetic Circuits with Nested Disjunctions“. Teoses: *Advances in Cryptology - CRYPTO 2021 - 41st Annual International Cryptology Conference, CRYPTO 2021, Virtual Event, August 16-20, 2021, Proceedings, Part IV*. Köide 12828. Lecture Notes in Computer Science. Springer, lk. 92–122. DOI: 10.1007/978-3-030-84259-8_4.

[109] Jens Groth. „On the Size of Pairing-Based Non-interactive Arguments“. Teoses: *Advances in Cryptology - EUROCRYPT 2016 - 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8-12, 2016, Proceedings, Part II*. Köide 9666. Lecture Notes in Computer Science. Springer, lk. 305–326. DOI: 10.1007/978-3-662-49896-5_11.

[110] Aniket Kate, Gregory M. Zaverucha ja Ian Goldberg. „Constant-Size Commitments to Polynomials and Their Applications“. Teoses: *Advances in Cryptology - ASIACRYPT 2010 - 16th International Conference on the Theory and Application of Cryptology and Information Security, Singapore, December 5-9, 2010. Proceedings*. Köide 6477. Lecture Notes in Computer Science. Springer, lk. 177–194. DOI: 10.1007/978-3-642-17373-8_11.

[111] Amos Fiat ja Adi Shamir. „How to Prove Yourself: Practical Solutions to Identification and Signature Problems“. Teoses: *Advances in Cryptology - CRYPTO '86, Santa Barbara, California, USA, 1986, Proceedings*. Köide 263. Lecture Notes in Computer Science. Springer, lk. 186–194. DOI: 10.1007/3-540-47721-7_12.

[112] Manuel Blum, Paul Feldman ja Silvio Micali. „Non-Interactive Zero-Knowledge and Its Applications (Extended Abstract)“. Teoses: *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, May 2-4, 1988, Chicago, Illinois, USA*. ACM, lk. 103–112. DOI: 10.1145/62212.62222.

[113] Ian Miers *et al*. „Zerocoin: Anonymous Distributed E-Cash from Bitcoin“. Teoses: *2013 IEEE Symposium on Security and Privacy*, lk. 397–411. DOI: 10.1109/SP.2013.34.

[114] Malte Möser *et al*. „An Empirical Analysis of Linkability in the Monero Blockchain“. *CoRR* (2017). URL: http://arxiv.org/abs/1704.04299.

[115] Reuters. *Explainer: 'Privacy coin' Monero offers near total anonymity*. 2019. URL: https://www.reuters.com/article/us-crypto-currencies-altcoins-explainer-idUSKCN1SL0F0.

[116] Vice. *The IRS Wants to Buy Tools to Trace Privacy-Focused Cryptocurrency Monero*. URL: https://www.vice.com/en/article/wxq9xx/the-irs-wants-to-buy-tools-to-trace-privacy-focused-cryptocurrency-monero.

[117] Shafi Goldwasser, Silvio Micali ja Charles Rackoff. „The Knowledge Complexity of Interactive Proof-Systems (Extended Abstract)“. Teoses: *Proceedings of the 17th Annual ACM Symposium on Theory of Computing, May 6-8, 1985, Providence, Rhode Island, USA*. ACM, lk. 291–304. DOI: 10.1145/22145.22178.

[118] Oded Goldreich, Silvio Micali ja Avi Wigderson. „Proofs that Yield Nothing But their Validity and a Methodology of Cryptographic Protocol Design (Extended Abstract)“. Teoses: *27th Annual Symposium on Foundations of Computer Science, Toronto, Canada, 27-29 October 1986*. IEEE Computer Society, lk. 174–187. DOI: 10.1109/SFCS.1986.47.

[119] Oded Goldreich, Silvio Micali ja Avi Wigderson. „Proofs that Yield Nothing But Their Validity for All Languages in NP Have Zero-Knowledge Proof Systems“. *J. ACM* 38.3 (1991), lk. 691–729. DOI: 10.1145/116825.116852.

[120] Claus-Peter Schnorr. „Efficient Identification and Signatures for Smart Cards“. Teoses: *Advances in Cryptology - CRYPTO '89, 9th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 1989, Proceedings*. Köide 435. Lecture Notes in Computer Science. Springer, lk. 239–252. DOI: 10.1007/0-387-34805-0_22.

[121] Ronald Cramer *et al*. „Multi-Autority Secret-Ballot Elections with Linear Work“. Teoses: *Advances in Cryptology - EUROCRYPT '96, International Conference on the Theory and Application of Cryptographic Techniques, Saragossa, Spain, May 12-16, 1996, Proceeding*. Köide 1070. Lecture Notes in Computer Science. Springer, lk. 72–83. DOI: 10.1007/3-540-68339-9_7.

[122] Ronald Cramer ja Victor Shoup. „A Practical Public Key Cryptosystem Provably Secure Against Adaptive Chosen Ciphertext Attack“. Teoses: *Advances in Cryptology - CRYPTO '98, 18th Annual International Cryptology Conference, Santa Barbara, California, USA, August 23-27, 1998, Proceedings*. Köide 1462. Lecture Notes in Computer Science. Springer, lk. 13–25. DOI: 10.1007/BFb0055717.

[123] Jens Groth ja Amit Sahai. „Efficient Non-interactive Proof Systems for Bilinear Groups“. Teoses: *Advances in Cryptology - EUROCRYPT 2008, 27th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Istanbul, Turkey, April 13-17, 2008. Proceedings*. Köide 4965. Lecture Notes in Computer Science. Springer, lk. 415–432. DOI: 10.1007/978-3-540-78967-3_24.

[124] *Signal*. URL: https://signal.org.

[125] Katriel Cohn-Gordon *et al*. „A Formal Security Analysis of the Signal Messaging Protocol“. Teoses: *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*. 2017, lk. 451–466. DOI: 10.1109/EuroSP.2017.27.

[126] Electronic Frontier Foundation. *Secure Messaging Scorecard 2020*. URL: https://www.eff.org/node/101713.

[127] The Guardian. *Signal: China appears to have blocked encrypted messaging app*. URL: https://www.theguardian.com/world/2021/mar/16/signal-blocked-china-encrypted-messaging-app.

[128] Nik Unger *et al.* „SoK: Secure Messaging". Teoses: *2015 IEEE Symposium on Security and Privacy*. 2015, lk. 232–249. DOI: 10.1109/SP.2015.22.

[129] Krishna Sampigethaya ja Radha Poovendran. „A Survey on Mix Networks and Their Secure Applications". *Proceedings of the IEEE* 94.12 (2006), lk. 2142–2181. DOI: 10.1109/JPROC.2006.889687.

[130] Valeh Farzaliyev, Jan Willemson ja Jaan Kristjan Kaasik. „Improved lattice-based mix-nets for electronic voting". *IET Information Security* 17.1 (2023), lk. 18–34. DOI: 10.1049/ise2.12089.

[131] Financial Times. *Estonia leads world in making digital voting a reality*. 2021. URL: https://www.ft.com/content/b4425338-6207-49a0-bbfb-6ae5460fc1c1.

[132] David L. Chaum. „Untraceable electronic mail, return addresses, and digital pseudonyms". *Communications of the ACM* 24.2 (1981), lk. 84–90. DOI: 10.1145/358549.358563.

[133] Jian Ren ja Jie Wu. „Survey on anonymous communications in computer networks". *Computer Communications* 33.4 (2010), lk. 420–431. DOI: 10.1016/j.comcom.2009.11.009.

[134] Roger Dingledine, Nick Mathewson ja Paul Syverson. *Tor: The second-generation onion router*. Tehniline raport. Naval Research Lab Washington DC, 2004. URL: https://apps.dtic.mil/sti/citations/ADA465464.

[135] Engadget. *Twitter launches a Tor service to help Russians evade censorship*. URL: https://www.engadget.com/twitter-tor-onion-service-evade-censorship-210549633.html.

[136] Wired. *How Tor Is Fighting — and Beating — Russian Censorship*. URL: https://www.wired.com/story/tor-browser-russia-blocks/.

[137] Emin Çalışkan, Tomáš Minárik ja Anna-Maria Osula. „Technical and legal overview of the tor anonymity network". *NATO Cooperative Cyber Defence Centre of Excellence* (2015). URL: https://ccdcoe.org/uploads/2018/10/TOR_Anonymity_Network.pdf.

[138] European Union Agency for Cybersecurity *et al. Privacy and data protection by design: from policy to engineering*. 2015. DOI: 10.2824/38623.

[139] Simone Fischer-Hübner, Luigi Lo Iacono ja Sebastian Möller. „Usable security und privacy". *Datenschutz und Datensicherheit-DuD* 34.11 (2010), lk. 773–782. DOI: 10.1007/s11623-010-0210-4.

[140] Marit Hansen. „Marrying transparency tools with user-controlled identity management". Teoses: *The Future of Identity in the Information Society: Proceedings of the Third IFIP WG 9.2, 9.6/11.6, 11.7/FIDIS International Summer School on The Future of Identity in the Information Society, Karlstad University, Sweden, August 4–10, 2007*. Springer, lk. 199–220. DOI: 10.1007/978-0-387-79026-8_14.

[141] Jonathan A. Obar ja Anne Oeldorf-Hirsch. „The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services". *Information, Communication & Society* 23.1 (2020), lk. 128–147. DOI: 10.1080/1369118X.2018.1486870.

[142] Centre for Information Policy Leadership. *Comments by the Centre for Information Policy Leadership on the Article 29 Data Protection Working Party's "Guidelines on Consent" adopted on 28 November 2017*. 2018. URL: https://www.huntonprivacyblog.com/wp-content/uploads/sites/28/2018/02/cipl_response_to_wp29_guidelines_on_consent-c.pdf.

[143] Shikun Zhang *et al*. „How Usable Are iOS App Privacy Labels?" *Proc. Priv. Enhancing Technol.* 2022 (4 2022), lk. 204–228. DOI: 10.56553/popets-2022-0106.

[144] Lorrie Faith Cranor. „Mobile-App Privacy Nutrition Labels Missing Key Ingredients for Success". *Commun. ACM* 65.11 (2022), lk. 26–28. ISSN: 0001-0782. DOI: 10.1145/3563967.

[145] Michele Nati ja Digital Catapult. *Personal Data Receipts: How transparency increases consumer trust. (2018)*. URL: https://www.digicatapult.org.uk/wp-content/uploads/2021/11/Personal_Data_Receipts_r1.5_2.pdf.

[146] Vitor Jesus ja Harshvardhan J. Pandit. „Consent Receipts for a Usable and Auditable Web of Personal Data". *IEEE Access* 10 (2022), lk. 28545–28563. DOI: 10.1109/ACCESS.2022.3157850.

[147] The White House. *US and UK launch innovation prize challenges in privacy-enhancing technologies to tackle financial crime and public health emergencies*. 2020. URL: https://www.whitehouse.gov/ostp/news-updates/2022/07/20/u-s-and-u-k-launch-innovation-prize-challenges-in-privacy-enhancing-technologies-to-tackle-financial-crime-and-public-health-emergencies.

[148] *H.R.1565 – Student Right to Know Before You Go Act of 2019*. 2019. URL: https://www.congress.gov/bill/116th-congress/house-bill/1565/text.

[149] Laura McKenna. *Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing*. Working Papers. URL: https://ideas.repec.org/p/cen/wpaper/18-47.html.

[150] Fair Lines America Foundation. *Declaration of John M. Abowd*. URL: https://s3.documentcloud.org/documents/21018464/fair-lines-america-foundation-july-26-2021-declaration-of-john-m-abowd.pdf.

[151] J. Tom Mueller ja Alexis R. Santos-Lozada. „The 2020 US Census differential privacy method introduces disproportionate discrepancies for rural and non-white populations". *Population Research and Policy Review* 41.4 (2022), lk. 1417–1430. DOI: 10.1007/s11113-022-09698-3.

[152] Alexis R. Santos-Lozada. „Changes in census data will affect our understanding of infant health". *Socius* 7 (2021). DOI: 10.1177/23780231211023642.

[153] The New York Times. *The 2020 Census Suggests That People Live Underwater. There's a Reason*. 2022. URL: https://www.nytimes.com/2022/04/21/us/census-data-privacy-concerns.html.

[154] The Washington Post. *New system to protect census data may compromise accuracy, some experts say*. 2021. URL: https://www.washingtonpost.com/local/social-issues/2020-census-differential-privacy-ipums/2021/06/01/6c94b46e-c30d-11eb-93f5-ee9558eecf4b_story.html.

[155] The Wall Street Journal. *Census Data Change to Protect Privacy Rattles Researchers, Minority Groups*. 2021. URL: https://www.wsj.com/articles/census-data-change-to-protect-privacy-rattles-researchers-minority-groups-11627902000.

[156] NPR. *For The U.S. Census, Keeping Your Data Anonymous And Useful Is A Tricky Balance*. 2021. URL: https://www.npr.org/2021/05/19/993247101/for-the-u-s-census-keeping-your-data-anonymous-and-useful-is-a-tricky-balance.

[157] Lincoln Journal Star. *Second resident of Nebraska's one-person town just a figment of Census Bureau's imagination*. 2021. URL: https://journalstar.com/news/state-and-regional/nebraska/second-resident-of-nebraska-s-one-person-town-just-a-figment-of-census-bureau-s/article_3dade4b8-8736-57af-9270-3a65973f5464.html.

[158] AP News. *Alabama drops lawsuit challenging Census privacy method*. 2021. URL: https://apnews.com/article/alabama-lawsuits-census-2020-redistricting-us-census-bureau-3c6f5eacc6c5638756700ba8308c45d2.

[159] Aloni Cohen *et al.* „Census TopDown: The Impacts of Differential Privacy on Redistricting". Teoses: Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. DOI: 10.4230/LIPICS.FORC.2021.5.

[160] Boston Women's Workforce Council. *The Gender and Racial Wage Gap Measurement in Boston by the Numbers*. URL: https://thebwwc.org/wage-gap-studies.

[161] Andrei Lapets *et al.* „Secure MPC for analytics as a web application". Teoses: *2016 IEEE Cybersecurity Development (SecDev)*. IEEE, 2016, lk. 73–74. DOI: 10.1109/SecDev.2016.027.

[162] BU Today. *Tackling the Wage Gap with Code*. 2017. URL: https://www.bu.edu/articles/2017/tackling-wage-gap-with-code.

[163] The Brink. *Using Data Science to Address the Gender and Racial Wage Gap*. 2021. URL: https://www.bu.edu/articles/2021/using-data-science-to-address-the-gender-and-racial-wage-gap.

[164] The Washington Post. *It's time to tell students what they need to know*. 2018. URL: https://www.washingtonpost.com/news/grade-point/wp/2018/08/21/its-time-to-tell-students-what-they-need-to-know.

[165] Ahmet Aktay *et al.* „Google COVID-19 Community Mobility Reports: Anonymization Process Description (version 1.0)". *CoRR* (2020). URL: https://arxiv.org/abs/2004.04145.

[166] Miguel Guevara. *Our latest updates on Fully Homomorphic Encryption*. 2021. URL: https://developers.googleblog.com/2021/06/our-latest-updates-on-fully-homomorphic-encryption.html.

[167] *What Are Privacy-Enhancing Technologies (PETs) and How Will They Apply to Ads?* 2021. URL: https://about.fb.com/news/2021/08/privacy-enhancing-technologies-and-ads.

[168] John Nguyen *et al.* „Federated Learning with Buffered Asynchronous Aggregation". *CoRR* (2021). URL: https://arxiv.org/abs/2106.06639.

[169] Matthias Paulik *et al.* „Federated Evaluation and Tuning for On-Device Personalization: System Design & Applications". *CoRR* (2021). URL: https://arxiv.org/abs/2102.08503.

[170] *Apple privacy features*. URL: https://www.apple.com/privacy/features/.

[171] Mark Russinovich *et al. CCF: A Framework for Building Confidential Verifiable Replicated Services*. Tehniline raport. Microsoft Research ja Microsoft Azure, 2019. URL: https://github.com/microsoft/CCF/blob/main/CCF-TECHNICAL-REPORT.pdf.

[172] Seth Patton. *Microsoft Viva Insights helps people nurture wellbeing and be their best*. 2021. URL: https://techcommunity.microsoft.com/t5/microsoft-viva-blog/microsoft-viva-insights-helps-people-nurture-wellbeing-and-be/bap/2107010.

[173] *Password Monitor: Safeguarding passwords in Microsoft Edge*. 2021. URL: https://www.microsoft.com/en-us/research/blog/password-monitor-safeguarding-passwords-in-microsoft-edge.

[174] NIST. *Roadmap for Advancing the NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management*. 2020. URL: https://www.nist.gov/system/files/documents/2020/01/16/NIST-Privacy-Framework-Roadmap-v1.0.pdf.

[175] Yoram Meijaard *et al*. *Netherlands Cryptoland. Starting point of the cryptocommunications roadmap: an overview of 4 important developements in cryptography*. 2021. URL: https://dcypher.nl/file/download/f67b5ad2-beee-4fdc-936c-5ac7b5fa3fea/netherlands-cryptoland-exploratory.pdf.

[176] *TNO's Early Research Programme in Next Generation Crypto*. 2022. URL: https://dcypher.nl/news/view/f3665636-5420-4151-9761-a6cbd80483e1/tnos-early-research-programme-in-next-generation-crypto.

[177] Miriam van der Sangen. *CBS explores possible privacy preserving techniques*. 2021. URL: https://www.cbs.nl/en-gb/corporate/2021/14/cbs-explores-possible-privacy-preserving-techniques.

[178] CBS. *Microdata: Conducting your own research*. URL: https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research.

[179] Erieke Weitenberg. *Secure and private statistics with distributed Paillier*. 2021. URL: https://medium.com/applied-mpc/secure-and-private-statistics-with-distributed-paillier-8a186410b5af.

[180] Technolution Spark. *Multi-Party Computation protects privacy-sensitive capacity data*. URL: https://www.technolution.com/spark/cases/multi-party-computation-protects-privacy-sensitive-capacity-data/?noredirect=en-GB.

[181] Ann Cavoukian. *Privacy by Design*. URL: https://web.archive.org/web/20110826032358/http://privacybydesign.ca/about.

[182] Ann Cavoukian *et al*. „Privacy and Video Surveillance in Mass Transit Systems: A Special Investigation Report" (2008). URL: https://www.ipc.on.ca/wp-content/uploads/2016/09/mc07-68-ttc_592396093750.pdf.

[183] Technology Analysis Division of the Office of the Privacy Commissioner of Canada. *Privacy Enhancing Technologies –- A Review of Tools and Techniques*. 2017. URL: https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2017/pet_201711.

[184] Brian J. Barth. *Death of a Smart City*. 2020. URL: https://onezero.medium.com/how-a-band-of-activists-and-one-tech-billionaire-beat-alphabets-smart-city-de19afb5d69e.

[185] Article 29 Data Protection Working Party. *Opinion 05/2014 on Anonymisation Techniques adopted on 10 April 2014*. 2014. URL: https://www.cnil.fr/sites/default/files/atoms/files/88197.pdf.

[186] VentureBeat. *France tries to salvage failed StopCovid tracing app as cases surge*. 2020. URL: https://venturebeat.com/mobile/france-tries-to-salvage-failed-stopcovid-tracing-app-as-cases-surge.

[187] Jean Ogier du Terrail *et al.* „Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer". *Nature medicine* 29.1 (2023), lk. 135–146. DOI: 10.1038/s41591-022-02155-w.

[188] Owkin, Inc. *First ever use of federated learning to train deep learning models on multiple hospitals, without data leaving firewalls*. 2023. URL: https://medicalxpress.com/news/2023-01-federated-deep-multiple-hospitals-firewalls.html.

[189] Nathalie Lassau *et al.* „Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients". *Nature communications* 12.1 (2021), lk. 1–11. DOI: 10.1038/s41467-020-20657-4.

[190] Arkhn. *PRESS RELEASE: Arkhn partners with Owkin and Inria to standardize access to health data from the IUCT-Oncopole, Institut Curie, Institut Bergonié and Toulouse University Hospital in a €10+ million project*. 2022. URL: https://arkhn.org/static/press_oncolab_june_2022_EN-7fb3722af4f59365439843fa01d1cbef.pdf.

[191] Infocomm Media Development Authority. *Privacy Enhancing Technologies (PET) Sandbox*. 2022. URL: https://www.imda.gov.sg/How-We-Can-Help/Data-Innovation/Privacy-Enhancing-Technologies-Sandbox.

[192] Hallam Stevens ja Monamie Bhadra Haines. „TraceTogether: Pandemic Response, Democracy, and Technology". *East Asian Science, Technology and Society: An International Journal* 14.3 (2020), lk. 523–532. DOI: 10.1215/18752160-8698301.

[193] Carmela Troncoso *et al.* „Deploying Decentralized, Privacy-Preserving Proximity Tracing". *Commun. ACM* 65.9 (2022), lk. 48–57. DOI: 10.1145/3524107.

[194] Centre for Data Ethics ja Innovation. *UK and US launch innovation prize challenges in privacy-enhancing technologies to tackle financial crime and public health emergencies*. 2022. URL: www.gov.uk/government/news/uk-and-us-launch-innovation-prize-challenges-in-privacy-enhancing-technologies-to-tackle-financial-crime-and-public-health-emergencies.

[195] LawtechUK. *Legal Data Access, Multi-party Proof of Concept*. 2022. URL: https://lawtechuk.io/explore/legal-data-access-multi-party-proof-of-concept.

[196] Uwe Serdult *et al.* „Fifteen years of internet voting in Switzerland [History, Governance and Use]". Teoses: *2015 Second International Conference on eDemocracy & eGovernment (ICEDEG)*. 2015, lk. 126–132. DOI: 10.1109/ICEDEG.2015.7114482.

[197] Swiss Post. *Swiss Post temporarily suspends its e-voting system*. 2019. URL: https://www.evoting-blog.ch/en/pages/2019/swiss-post-temporarily-suspends-its-e-voting-system.

[198] Swiss Post. *Public intrusion test: ethical hackers can attack the Swiss Post e-voting infrastructure*. 2022. URL: https://www.evoting-blog.ch/en/pages/2022/public-intrusion-test-ethical-hackers-can-attack-the-swiss-post-e-voting-infrastructure.

[199] Swiss Post. *Cryptographic Primitives of the Swiss Post Voting System, Pseudo-code Specification, Version 1.2.1*. URL: https://gitlab.com/swisspost-evoting/crypto-primitives/crypto-primitives/-/blob/master/Crypto-Primitives-Specification.pdf.

[200] Carmela Troncoso *et al.* „Deploying Decentralized, Privacy-Preserving Proximity Tracing“. *Commun. ACM* 65.9 (2022), lk. 48–57. DOI: 10.1145/3524107.

[201] Jean Louis Raisaro *et al.* „MedCo: Enabling Secure and Privacy-Preserving Exploration of Distributed Clinical and Genomic Data“. *IEEE/ACM transactions on computational biology and bioinformatics* 16.4 (2018), lk. 1328–1341. DOI: 10.1109/TCBB.2018.2854776.