TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Anett Pärismaa

Eestikeelsete paroolide mustrite uurimine ja ründesõnatiku koostamine

Bakalaureusetöö (9 EAP)

Juhendaja: Alo Peets

Tartu 2024

Eestikeelsete paroolide mustrite uurimine ja ründesõnastiku koostamine

Lühikokkuvõte:

Enamik inimesi kasutab igapäevaselt autentimiseks mitmeid paroole, mistõttu on oluline nende kohta rohkem teada saada. Siinse töö eesmärk on uurida eestikeelsete kasutajate paroolide loomise harjumusi ning mustreid. Mainitut arvesse võttes antakse töös soovitusi äraarvatavate paroolide vältimiseks ja turvapraktika tugevdamiseks. Töö käigus koguti Internetis lekkinud eesti kasutajate paroole, samuti viidi inimeste paroolide loomise harjumuse kaardistamiseks läbi küsitlusuuring. Tulemuste rakendamiseks loodi veebisait, kus saab testida, kas otsitav parool leidub umbes 50 miljoni kirjega näidis-ründesõnastikus. Koostatud ründesõnastiku kogumaht on üle 50 GB. Turvatestimisel saab selle abil tuvastada haavatavaid paroole ja mustreid. Kirjalikus osas tutvustatakse küsitlusuuringu, veebirakenduse ning eestikeelse ründesõnastiku koostamise protsessi ja analüüsi tulemusi.

Võtmesõnad: parool, salasõna, turvatestimine, sõnastikrünne, eestikeelne ründesõnastik

CERCS: P175 Informaatika, süsteemiteooria

Researching the Patterns of Estonian Language Passwords and Compiling a Password Dictionary

Abstract:

People use a variety of passwords for authentication on a daily basis, hence it is crucial to gain more insight into them. The aim of this study is to investigate the password creation habits and patterns of Estonian-speaking users. This thesis provides recommendations for avoiding predictable passwords and strengthening security practices. Passwords of Estonian users that had leaked online were collected, and a survey was conducted to map out people's password creation habits. To implement the findings, a website was developed where one can test whether a password can be found in a sample attack dictionary that has approximately 50 million entries. The compiled attack dictionary has a total volume of over 50 GB. This can be used in security testing to identify vulnerable passwords and patterns. The written part introduces the survey research, web application, and the process of compiling the Estonian attack dictionary and the analysis results.

Key words: password, secret word, penetration testing, dictionary attack, password dictionary in Estonian

CERCS: P175 Informatics, systems theory

Sisukord

Sissejuhatus 5

1. Terminid 7

2. Ülevaade tehtud töödest 10

3. Paroolidega seotud ründed ja ennetustegevus 12

3.1 Jõurünne 12

3.2 Sõnastikrünne 13

3.3 Õngitsusrünne 14

3.4 Rünnete ennetamine E-ITS'i näitel 15

4. Eestikeelsete paroolide kogumine 17

4.1 Küsimustiku koostamise taust 17

4.2 Küsimustiku loomise alused 18

4.3 Andmete kogumine Internetist 20

4.4 Andmete kogumine veebilehelt 22

5. Veebilehe arendus 23

5.1 Funktsionaalsed ja mittefunktsionaalsed nõuded 23

5.1.1 Funktsionaalsed nõuded 23

5.1.2 Mittefunktsionaalsed nõuded 24

5.2 Arhitektuur 24

5.3 Tehnoloogiad 25

5.3.1 Konteineritehnoloogia 25

5.3.2 Veebirakendus 26

5.3.3 Andmebaas 27

5.3.4 Domeeni registreerimine ja serveri seadistamine 28

5.4 Testid 28
5.5 Disain 29
5.6 Salafraaside generaator 29
6. Analüüs 30
6.1 Küsitlusuuring 30
6.1.1 Harjumused 30
6.1.2 Paroolide genereerimine 31
6.1.3 Stsenaariumid 35
6.1.4 Demograafiline teave 36
6.2 Lekkinud paroolid 36
6.3 Veebilehe tulemused 37
7. Tulemused 39
7.1 Sõnastiku koostamine 39
7.2 John the Ripper'i kasutamine 40
7.2.1 Sõnastikrünne 41
7.2.2 Maskeerimine 42
7.2.3 Soolatud räsi murdmine 43
Kokkuvõte 45
Kasutatud allikad 46
Lisa 1 49
Litsents 50

Sissejuhatus

Pea kõik inimesed kasutavad igapäevaseks autentimiseks palju erilaadi paroole. Aktiivse arvuti- ja Interneti-kasutaja poolt vajatavate salasõnade hulk võib küündida mitmekümneni või isegi üle selle. Keeruline ja tülikas on neid kõiki meeles pidada, mistõttu hakatakse tarvitama lihtsaid paroole. Neid kiputakse taaskasutama või järgitakse levinud mustreid nagu suurtäht alguses ja number lõpus. See muudab kasutajakontod haavatavaks ja seab ohtu kasutajate isikliku info.

RIA (Riigi Inosüsteemi Amet) 2024. aasta küberturvalisuse aastaraamatu [1] andmetel on vaatamata teadlikkuse kasvule endiselt murekohaks paroolide ristkasutamine (töö- ja erakontol), mis muudab kontod haavatavamaks ning annab võimaluse meilipette rünneteks. Võib täheldada, et inimeste küberteadlikkus ei pea sammu kurjategijate meetoditega ning tugevad autentimismeetmed pole veel piisavalt laialdaselt levinud, et kontosid kaitsta.

Kristjan Pühveli 2022. aasta bakalaureusetöö [2] tulemusena õnnestus märkimisväärne hulk Tartu Ülikooli parooliräsidest lahti murda, mistõttu ülikooli turvameetmeid parandati ja kontodele kehtestati kaheastmeline autentimine. Eestikeelset ründesõnastikku kasutades saadi 1476 räsist lahti 1284. Kuna K. Pühveli töö ei ole avalik, siis on praeguse uurimuse eesmärk avada teematikat nii, et laiem ringkond saaks paroolidega seotud probleemidega tutvuda ning tulemuste põhjal oleks võimalik pakkuda soovitusi turvameetmete parandamiseks.

Paroolide murdmise efektiivsuse puhul on oluline kultuuriliste omapärade arvestamine [3]. Kurjategijate kõrval peaks seda arvesse võtma ka infoturbejuhid ja turvaspetsialistid, et süsteeme selliste rünnete eest kaitsta. Eestis tegeleb RIA küll inimeste ja ettevõtete teadlikkuse tõstmisega ja meetmete loomisega kübervaldkonnas (nt E-ITS[1]), kuid konkreetselt eestikeelseete paroolide mustreid analüüsivad teadustööd või lahendusi, nt haavatavate paroolide sõnastik, ei õnnestunud Internetist leida. Kasutatud otsingusõnad olid *paroolide sõnastik*, *eestikeelsed paroolid*, *estonian password list*.

Töö eesmärk on pakkuda lahendusi paroolidega seotud rünnete ennetamiseks. Eestikeelsete kasutajate paroolide loomise harjumusi ja mustreid analüüsides on võimalik jagada soovitusi turvapraktikate parandamiseks. Käesoleva töö üheks väljundiks on veebirakendus, kus kasutatakse genereeritud eestikeelsete paroolide sõnastikku, mida testitakse kasutaja sisendi vastu.

[1] Eesti Infoturbestandard. https://eits.ria.ee

Töö koosneb seitsmest osast, mis on jaotatud alapeatükkidesse. Ülevaade antakse terminitest, tehtud töödest, rünnete tüüpidest, andmete kogumise metoodikast (küsitlusuuring, andmelekked ning veebileht), rakenduse Paroolikratt[2] loomisest, andmete analüüsimisest ning viimaks tulemustest (sõnastiku koostamine ja John the Ripperi kasutamine), sellele järgneb kokkuvõte. Töö käigus loodud failid ja materjalid paiknevad lisades.

[2] Paroolikratt. https://paroolikratt.ee

1. Terminid

Töö üheseks mõistmiseks ning tulemuste tõlgendamiseks on vajalik anda ülevaade kasutatavatest terminitest. Järgnevalt on selgitatud peamised terminid ning avatud kasutamise konteksti. Vajadusel on terminid eestindatud. Lugejale, kes on valdkonnaga kursis, on esitatud informatsioon arvatavasti juba tuttav ning sel juhul võib töö lugemist järgnevast peatükist jätkata.

Autentimine on protsess, mis kontrollib esitatud identiteeti. Viimane koosneb tavaliste rakenduste puhul kahest osas –kasutajanimest ja paroolist.

Kasutajanimi (ingl *user ID*) *on kordumatu märgijada või kujund, millega andmetöötlussüsteem identifitseerib kasutajat.*[4] Siinpuhul manitseb AKIT kasutajanime mitte kasutajatunnusega segamini ajama. Kasutajatunnus on atribuut, milleks võib olla mõni element või biomeetrik.[4]

Parool (ingl *password*) on *konfidentsiaalne märgijada autentimiseks kasutatava identsustõendina pääsu reguleerimise mehhanismides.*[4] Selle võib kergesti segamini ajada eesti keeles samatähenduslikult kasutatava **salasõnaga** (ingl *secret word*), mis erialaspetsiifiliselt tähendab *märgijada, mis on peidetud tekstilisse robotilõksu ja tuleb edasipääsuks sisestada*.[4] Samasse komplekti kuulub veel paroolfraas või **salafraas** (ingl *passphrase*), mis kätkeb endas salajast lauset, mille eesmärk on hõlbustada pika parooli meeldejätmist. Neid termineid kasutatakse selles töös sünonüümidena (parooli definitsioonis), sest keeleliselt on need juurdunud ja vastupidine praktika võiks tekitada segadust. Ühe uuema terminina esineb *passkey*, mida võiks tõlkida kui **salavõti**. See tekkis uue tehnoloogiana vastukaaluks salasõnade ning paroolide meeleshoidmisele. Salavõti kujutab endast krüptograafilist seadmepõhist identsustõendit, millega pääseb rakendustele ligi ning see on väidetavalt õngitususründekindel ettevõtte FIDO[3] (Fast IDentity Online) tehnoloogia.

Küberkaitse (ingl *cybersecurity, cyber defence*) eesmärk on kaitsta teenuseid, süsteeme ning laiemalt digitaalset vara küberrünnete eest, kuid terminina on see ühtselt määratlemata.[4] Selle töö raames kätkeb küberkaitse endas küberturvet, infoturvet ja võrguturvet. Ehk kaitse on protsess mingi välise ründe vastu ning turve on sisemise oleku säilitamine või parandamine turbetegevusi teostades.

[3] FIDO Alliance. https://fidoalliance.org
[4] Andmekaitse ja infoturbe portaal. Cybernetica AS. https://akit.cyber.ee

Tasub välja tuua, et küberkaitse kontekstis ei ole **rünne** sama, mis **rünnak**. Kuigi Sõnaveebis[5] käsitletakse neid sünonüümidena ka spetsiifiliselt küberkaitse terminite puhul, näiteks teenusetõkestusrünne ja teenusetõkestusrünnak. Andmekaitse ja infoturbe leksikon (AKIT) hoiatab nende sõnade sünonüümina kasutamise eest.[6] Rünnak on mõlemas kontekstis *äge ja vägivaldne, sageli relvastatud kallaletung kellelegi või millelegi; sõjaline pealetung, kus hrl tulistatakse kõikidest relvadest.*[5] Samas, kui rünne on küberkaitses *katse infosüsteemi ja/või selles olevat teavet hävitada, muuta, paljastada, muuta või blokeerida või muul viisil rikkuda turvapoliitikat.*[5] Täpsemalt, küberrünne (ingl *cyberattack*) on *rünne, mis toimub küberruumi kaudu ja on suunatud küberruumi kasutamisele, püüdes*.[6]

- häirida, pärssida, hävitada või kahjustavalt valitseda andmetöötluse keskkonda või taristut või

- rikkuda andmete terviklust või

- varastada tundlikku teavet.

Segadust tekitavaid sõnapaare on veelgi, näiteks **internet** ja **Internet**. Väikese algustähega variant on *arvutivõrkude vaheline ühendus, mis võimaldab mitut tüüpi sideteenuseid (nt sõnumsidet, failiedastust); mis tahes arvutivõrkude üldnimetus*.[5] Suurtähega algav Internet on *kokkuühendatud avalike võrkude ülemaailmne süsteem*.[5] Sõnaveebis on veel näitena toodud, et e*smalt täitis interneti rolli Darpanet, hiljem Internet*.[5] Mis võiks anda indikatsiooni nende terminite kasutamisest. Selles töös kasutatakse internetti üldnimetusena ja Internetti avalike võrkude süsteemina.

Combo list *on tekstifail, mis koosneb lekkinud kasutajanimede ja salasõnade paaridest* [4]. Need võivad olla esitatud lihttekstina, selliseid faile kasutatakse sageli rünnete teostamisel.

Base64 on *kodeerimisüsteem, mis teisendab suvalise baidijada ASCII kirjamärkideks 64-märgilises alamhulgas (inglise suur- ja väiketähed, kümnendnumbrid, plussmärk, kaldkriips).*[6] Nt, „parool“ puhul on väljund *cGFyb29s*. Lisades tühiku, siis „parool “ puhul *cGFyb29sIA==*.

Räsi on *püsipikkusega sõnumilühend, mis on andmetest saadud räsifunktsiooni abil, ühesuunalise krüpteerimisega, st pöördumatu tihendusega*.[6] Räsimine on räsifunktsiooni, nt SHA (*Secure Hash Algorithm*), kasutav protsess, mille tulemusena saavutatakse räsi.

[5] Sõnaveeb. https://sonaveeb.ee
[6] Andmekaitse ja infoturbe portaal. Cybernetica AS. https://akit.cyber.ee

Ründesõnastik on potentsiaalsetest paroolikandidaatidest koosnev sõnade loend, mida kasutatakse sõnastikründe teostamisel. See võib olla ette arvutatud, koosneda sõnakombinatsioonidest või lihtkujul sõnadest, mida ründe käigus modifitseeritakse.[7]

Sool on *krüpteerimisel või räsimisel lisatav mittesalajane (tavaliselt juhuarvuline) väärtus, näiteks parooli krüpteerimisel talle lisatav täidis, mis raskendab sõnastikrünnet*.[7] Soolamine on protsess, mille käigus lisatakse paroolile sool ja seejärel toimub krüpteerimine või räsimine.

Sellega on antud ülevaade töös kasutatavatest spetsiifilisematest terminitest. Järgnevate peatükkide lugemisel saab vajaduse korral siia tagasi pöörduda.

[7] Andmekaitse ja infoturbe portaal. Cybernetica AS. https://akit.cyber.ee

2. Ülevaade sarnastest töödest

Lam Tran jt [5] on koostanud ülevaate aastatel 1979 – 2022 ilmunud uurimustest, mis käsitlevad salasõnade murdmise tehnoloogiaid. Erinevad praktikad kätkevad endas nii juhuslikku arvamist kui ka masinõppe meetodeid. Tuuakse välja, et rünnakumudeli koostamine ei hõlma ainult lekkinud paroolide rakendamist. Kasulik on integreerida ka muid lekkinud isikuandmeid. Samuti tõstab efektiivsust see, kui kasutada ära keele omapärasid ning klaviatuuripaigutust. Autorid annavad soovituse arendada paroolide murdmise uurimist edasi, sest vaid nii saab tõsta üldist paroolide turvalisuse taset.

Spetsiifilist paroolide murdmise arendustööd on ka jätkatud. Hazel Murray ja David Malone [3] koostasid mudeli, mis põhineb MAB (*multi-armed bandit*). Töös tuuakse välja, et loodud mudel õpib kiiresti. Arvates ära ühe kasutaja parooli, saab see potentsiaalselt hakkama ka teiste sarnase grupi paroolidega. Seetõttu annavad autorid soovituse koostada veebilehtede haldajatel keelatud paroolide nimekirjad, et vältida salasõnades erialaseptsiifiliste mustrite kordumist. Samuti tasub välja tuua, et Murray ja Malone koostasid erinevate näitajate (nt allikas, keel, rahvus) põhjal eraldi sõnastikud. See osutus palju efektiivsemaks ja dünaamilisemaks viisiks paroolide murdmisel kui varem kasutatud klassifitseerimata sõnastikud. Seega saab väljatoodud tööde põhjal järeldada, et valitud valdkonnas on puudu töödest ning uurimustest, mis arvestavad keelelist ja kultuurilist konteksti.

Tartu Ülikoolis proovis Kristjan Pühvel 2022. aastal bakalaureusetöö [2] töö raames murda ca 19 000 TÜ kasutajakonto paroole. Selleks kasutas ta omaloodud tarkvara ning sõnastikke, tänu millele õnnestus 1476 NTLM räsi ära arvata. Inglisekeelse ründesõnastikuga saadi lahti 192 ja eestikeelsega 1284 räsi. Murdmine teostati n-ö laborikeskkonnas, sest TÜ IT-osakonna poolt käivitati skript juba kasutajakontode ja vastavate parooliräside andmebaasis nii, et algoritm ei pidanud arvestama süsteemipiirangutega ning töö autoril puudus võimalus kasutaja tema parooliga kokku viia. Tulemustest avaldus ka reaalne mõju, sest nüüdsest kasutatakse ülikoolis mitmeastmelist autentimist. Tulemusi esitleti lühidalt konverentsil „Küberinnovatsioon 2022“ [6]. Töö ise on avaldamispiiranguga kuni 10.05.2025 [2].

Kuigi mujalt maailmast leiab uurimusi, kus keskendutakse salasõnade mustrite analüüsimisele, siis analoogilisi eestikeelseid töid leida ei õnnestunud. Teadustööde otsingumootorites kasutati märksõnu *password authentication*, *password characteristics parool*, *password cracking*, *survey*

on estonian passwords. He jt [7] uurisid levinud salasõnu, nende struktuuri ja tähtede jaotumist. Põhifookus oli erinevate gruppide võrdlemisel. Vaatluse all olid Hiinas nelja kohalikku provintsi hõlmav rongiliikluse veebisaiti, inglisekeelsetest keskkondades saadi Fairwritheri ning Facebooki andmed. Analüüsi tulemusena leiti, et gruppidevahelised geograafilised erinevused (sh keelelised) põhjustavad märgatavaid erinevusi uuritud indikaatorite (erinevate märkide sagedus ja paiknemine) osas. Samuti täheldati erinevuseid paroolide tugevuses provintside vahel.

Viktor Taneski jt [8] selgitasid küsitlusuuringu teel välja paroolide loomise mustreid, kus koguti andmeid turismiteaduskonna tudengitelt ning arvutitehnika ja informaatika üliõpilastelt. Eesmärk oli võrrelda inimeste küsitluste vastuseid reaalsete paroolidega. Selgus, et turismitudengite antud vastused ja reaalsed paroolid kattusid. Samas tehnoloogiavaldkonna õppurite puhul esines oluline erinevus päris salasõnade ja küsimustiku vastuste vahel. Informaatika ja arvutitehnika üliõpilaste paroolid olid üldiselt ka nõrgemad kui turismitudengite omad.

Peatükis anti mõningane ülevaade valdkonnas tehtud töödest, mis pole kaugeltki täielik, sest valdkond on lai ning uurimisallikaid palju. Paroolide puhul tuleb arvestada keelelist ja kultuurilist konteksti. Käesolevale tööle samalaadse lähenemisega uurimust leida ei õnnestunud. Järgmises peatükis tutvustakse üldisemalt rünnete tüüpe ning seejärel suundutakse tööprotsessi ning tulemuste juurde.

3. Paroolidega seotud ründed ja ennetustegevus

Paroolidega kaitstud keskkonnad tekitavad huvi küberkurjategijates, mistõttu võivad nad andmete omastamise ja tulu teenimise eesmärgil IT-süsteemide vastu ründeid korraldada. Leidub erinevaid viise, kuidas paroole teada saada. Selles töös piirdutakse lühiülevaate andmisega jõurünnetest (ingl *brute-force attack*), mille üheks näiteks on omakorda sõnastikrünne (ingl *dictionary attack*). Samuti käsitletakse pealiskaudselt suhtlusrünnet (ingl *social engineering*), mis sageli lõimub teiste ründeviisidega, sest inimesega manipuleerimine ning psühholoogiline mõjutamine on tihti odavam ja lihtsamini teostatavam kui kallist riistvara ning tehnilisi oskusi nõudvad ründed. Järgnevates alampeatükkides selgitatakse, mis on jõurünne ning selle erinevad tüübid. Täpsemalt analüüsitakse sõnastikründe aspekte.

3.1 Jõurünne

Jõurünne on oma olemuselt *parooli, krüptovõtme vm salajase mõistatamine kõiki võimalikke väärtusi läbi proovides*.[8] Ründajad reaaliseerivad seda katsetades kõikvõimalikke kasutajanimede ning paroolide kombinatsioone. Selle miinused on suur aja- ja energiakulu ning ebaefektiivsus. Tänapäeval on paljud süsteemid kõige algelisemate jõurünnete vastu kaitstud mitmeastmelise autentimise (ingl *multi-factor authentication*) ja sisestuskordade piiramise kaudu. Veel on levinud erinevad biomeetrilised lahendused (sõrmejälg, näotuvastus), sertifikaadi abil autentimine, tookenid (ingl *token*) ja salavõtmed [9]. Samuti on üheks meetmeks paroolidele seatud nõuded, nt erisümbolid, mis tõstavad kombinatsioonide arvu. Näiteks on inglise tähestikus oleva 26 väiketähe hulgast 5-märgilise parooli moodustamiseks (märkide järjekord on oluline ning nende hulk ei vähene) $26^5 = 11\,881\,376$ võimalust. Sama pikkusega parooli moodustamiseks eesti keele tähestikus (koos võõrtähtedega) tõuseb võimalike 5-märgiliste paroolide arv $32^5 = 33\,554\,432$ni. Seega pea kolmekordistub. Täna on sobiliku riistvara ja arvutusvõimsuse juures võimaluste hulk marginaalne, kuid jõudluse võimekust alandavad süsteemipiirangud.

Jõurünne jaguneb veel [10]:

- Sõnastikründeks (ingl *dictionary attack*), mis tähendab, et ründaja kasutab ründe teostamiseks teatud tunnuste alusel kokku pandud sõnade ja fraaside kogumit;

[8] Andmekaitse ja infoturbe portaal. Cybernetica AS. https://akit.cyber.ee

- Tabeliründeks (ingl *rainbow table attack*), mis kasutab eelarvutatud räsidega (ingl *hash*) tabelit;
- Kontovarguseks (ingl *credential stuffing*), mille puhul ründaja eeldab, et kasutaja kordab samasid kasutajakonto andmeid erinevate keskkondade puhul;
- Tavaline jõurünne (ingl *simple brute-force attack*), kus ründaja eesmärk on proovimise teel parool ära arvata;
- Pööratud jõurünne (ingl *reverse brute-force attack*), mille puhul ründaja teab parooli ning proovib kasutajanime ära arvata.

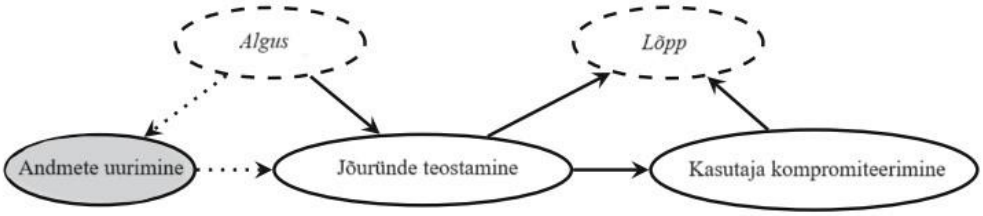
OWASP[9] (Open Worldwide Application Security Project) on mittetulundusühing, mis tegeleb turvalise tarkvara loomise propageerimisega. Nad on koostanud standardi juhtimaks tähelepanu enimlevinud tarkvara turvalisuse vigadele ehk OWASP Top 10 [11]. A07 on *Identification and Authentication Failures*[10] ehk identifitseerimine ja autentimise vead, mis kätkeb endas muuhulgas jõurünnet. Ründe ennetamiseks pakutakse lisaks välja ka paroolide kontrollimist enimlevinud salasõnade vastu (halvimate paroolide top 10 000), vaikeparoolide kasutamise vältimist ning veebiühenduse (sessiooni ID) turvalisemaks muutmist.

3.2 Sõnastikrünne

Sõnastikrünne (ingl *dictionary attack*) on jõuründe liik, mis (enamasti) paroolide dekrüpteerimiseks kasutab levinud sõnadest ning sümbolitest koostatud tabelit, kus sõnakombinatsioonid võivad olla ettearvutatud.[3]

Hofstede jt [12] uurimuses on välja toodud veebipõhise sõnastikründe erinevad etapid, kuid seda võib laiendada ka üldisele ründe toimimisele.

[9] OWASP. https://owasp.org
[10] A07:2021 – Identification and Authentication Failures. https://owasp.org/Top10/A07_2021-Identification_and_Authentication_Failures



Joonis 1. Sõnastikründe etapid, kohandatud *[12]*.

Joonise 1 järgi jaguneb rünne jaguneb enamasti kolmeks osaks, milleks on:

- süsteemi läbivaatuse faas, kui ründaja otsib sihtmärki;
- ründefaas, kui teostatakse konkreetne rünne (sõnastikründepuhul proovitakse kasutajanime ning parooli kombinatsiooni);
- kompromiteerimise faas, kui toimub turvapoliitika rikkumine ja vara omastamine, mida võidakse kohe või mõne aja pärast ära kasutada.

Sõnastikrünne on märgatavalt efektiivsem kui tavaline jõurünne. Bošnjaki jt [13] tehtud töös tuleb välja, et MD5-ga räsitud paroolide puhul kulub tavaarvutil jõuründe korral 10 h 37 min, kuid sõnastikrünne võtab vaid kuus sekundit. SHA-1 puhul kasvab vahe veelgi, 21 h 43 min sõnastiku seitsme sekundi vastu. Nende näites töötas see kõige paremini inglise keelt kõnelevate kasutajate hulgas.

3.3 Õngitsusrünne

Õngitsusrünne on moodus, mille abil võib ründaja teada saada tundlikke andmeid, nt paroolid, kasutajanimed. Õngitsus on *teesklus, mille sooritaja saadab tundliku teabe saamiseks sõnumeid, mis näivad tulevat usaldatavast allikast, näiteks sotsiaalvõrgust, oksjonisaidist, pangast.*[11] Näiteks saadab ründaja libameilid, kus palutakse kasutajal oma parool vahetada, suunates kasutaja ründaja poolt hallatavale veebisaidile, kus ohvri kasutaja ning salasõna salvestatakse. See on üks võimalus, kuidas paroolid võivad lekkida.

[11] Andmekaitse ja infoturbe portaal. Cybernetica AS. https://akit.cyber.ee

3.4 Rünnete ennetamine E-ITS'i näitel

E-ITS[12] (Eesti Infoturbestandard) on RIA (Riigi Infosüsteemi Amet) poolt välja töötatud dokument, mis sisaldab meetmeid ning nõudeid IT-süsteemide ja andmete kaitseks. E-ITS on kohustuslik ca 3500 organisatsioonile, sealhulgas RIA ise [1].

Eesti Infoturbestandardi [14] peadokument on etalonturbe kataloog, mida täiendavad veel näiteks alusohtude kataloog, auditeerimis-, rakendus- ja riskihaldusjuhendid ning rollisõnstik. Sõna *parool* on alusohtude kataloogis[13] märgitud 21 korral, seda on mainitud üheteistkümne ohu kirjelduses, nt spionaaž, pealtkuulamine, kahjurprogrammid. Psühholoogilise manipuleerimise all on öeldud, *et paljud kasutajad kasutavad küll tugevat parooli, kuid sama parooli mitme kontoga*,[13] mida rõhutati ka RIA küberturvalisuse aastaraamatus [1]. Etalonturbe kataloogis [14] organisatsiooni ja personali moodulis (ORP.2.3) hoiatatakse paroolide ristkasutamise eest. IT-süsteemide moodulis (SYS.1.1 Oht 2.4) rõhutatakse eeliskonto kaitsmist parooliräside lekkimise ja jõuründe eest. Etalonturbe kataloog pakub ka meetmeid ohtudega ennetamiseks. Organisatsiooni ja personali moodulis (ORP.4.M22) juhitakse tähelepanu paroolide kvaliteedinõuete kehtestamisele:

a. *Nõuded parooli kvaliteedile on kehtestatud arvestades IT-süsteemide kaitsetarvet ja kasutajate profiili.*
b. *Parool peab olema piisavalt keerukas, et parooli ei saaks ära arvata*.
c. *Soovitatav on kasutada vähemalt 12 tähemärgi pikkuseid paroole.*
d. *Parool ei tohi olla nii keeruline, et kasutaja ei suuda seda regulaarse kasutuse puhul mõistlike pingutustega meelde jätta.*

Rakenduste moodulis (APP.2.2 Lisa 4.2) esitatakse rühmapoliitika turvasätete näidis, kus juhitakse muuhulgas tähelepanu paroolide maksimaalsele ja minimaalsele vanusele, ajaloo ning keerukuse nõuetele.

IT-süsteemi standardiga vastavusse viimine on pikaajaline protsess, mille käigus peab kasutajaid korduvalt harima ning teostama süsteemis regulaarseid turvakontrolle, nt paroolide testimine andmebaasis. Sellele vaatamata on meetmete rakendamine äriliselt ja majanduslikult kasulik, sest aitab vältida nii rahalist kui ka mainekahju.

[12] Eesti Infoturbestandard. https://eits.ria.ee
[13] Alusohtude kataloog. Eesti Infoturbestandard. https://eits.ria.ee/et/abimaterjalid/alusohtude-kataloog

Internetis leidub mitmeid soovitusi turvaliste salasõnade koostamiseks. Näiteks on RIA lehel [15] paroolide loomise ja kasutamise nõuannete all välja toodud nimekiri Eesti kasutajate kahekümnest levinumast paroolist (*123456*, *parool*, *qwerty* jne). Arvestades üha rangemaks muutuvaid nõudeid, on väheusutav, et 2024. aastal lubaksid veebikeskkonnad kasutajal sellesarnaseid paroole määrata. Ei saa välisada, et neid siiski leidub, aga enamasti nõutakse kasutajalt parooli koostamisel väike- ja suurtähtede, numbrite ning erisümbolite sisaldumist. Levitades teadmist, et nõrk parool on stiilis *12345* luuakse pettekujutelm, et nt *Parool1!* on tugev, sest vastab nõuetele. Tegelikult kattub loodud salasõna tüüpilise parooli mustriga ning võib seetõttu olla haavatav. Seega on paroolide mustritele tähelepanu pööramine kasutaja ja IT-süsteemi vaatest vägagi oluline.

4. Eestikeelsete paroolide kogumine

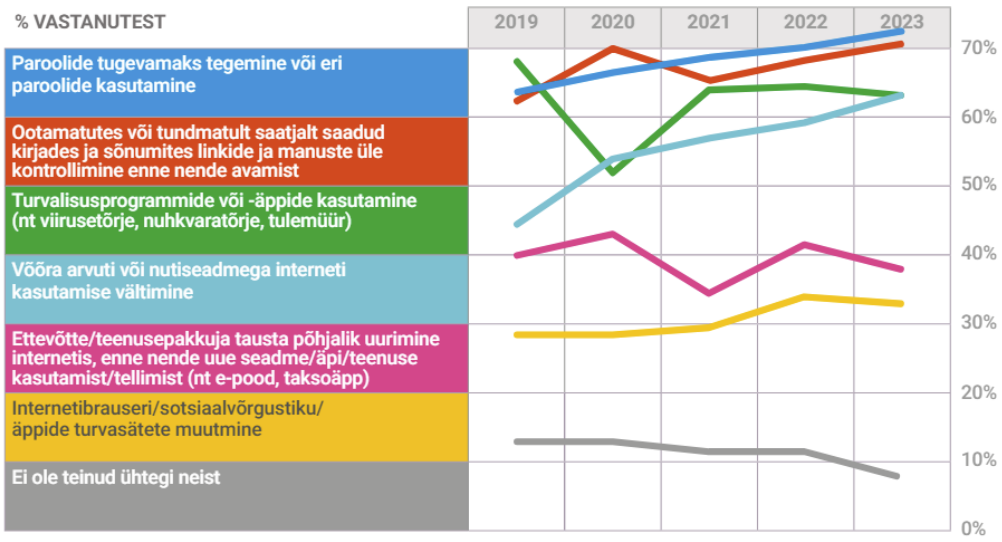
Töö eesmärgiks olnud sõnastiku koostamiseks koguti andmeid nii pinnaveebist kui ka küsimustiku vormis praegustelt Interneti-kasutajatelt. Andmelekete puhul on enamasti tegu aegunud infoga, mis ei pruugi kajastada tänaste veebikeskkondade turvanõudeid, nt RIA lehel [15] välja toodud levinud paroolid (*123456789, lammas, minaise* jne). Ankeedist saadud vastused kirjeldavad praeguste kasutajate harjumusi ning mustreid. Järgnevalt antakse ülevaade koostatud küsitlusuuringust ning tutvustatakse Internetist leitud avalikke lekkinud paroolide andmekogusid.

4.1 Küsimustiku koostamise taust

Vastavalt Murray ja Malone'i tehtud uurimusele on paroolide koostamisel oluliseks teguriks kultuurilised ja keelelised eripärad [3]. Vaatamata inimeste teadlikkuse ja veebikeskkonna turvalisuse kasvule, on paroolidega seotud riskid endiselt relevantsed. Ettevõtte IBM 2023. aasta aruanne [16] on tehtud 16 riigis toimunud 553 andmelekke kohta, jättes välja väga suured või marginaalsed intsidendid. Nende uuringust selgus, et maailmas oli andmelekke keskmine kulu ettevõttele 4,45 miljonit, USAs oli see keskmiselt 9,48 miljonit dollarit . IT-süsteemi turvalisusesse panustavatel ettevõtetel õnnestus andmelekke puhul säästa 1,68 miljonit dollarit võrreldes nendega, kes tegid seda vähesel määral või üldse mitte.

Kristjan Pühvel teostas 2022. aastal oma bakalaureusetöö [6] raames TÜ kasutajakontode turvatestimise, mille käigus õnnestus 19 000 aktiivsest kontost lahti murda 1496 kasutaja parool. Neist 192 inglisekeelset ning enda genereeritud sõnastikku kasutades 1304 eestikeelset. 20% murtud parooliga kontoomanikest vahetas ebaturvalise salasõna nädala jooksul peale vastavasisulise meili saamist. Tulemustest võib järeldada vajakajäämisi kasutajate küberteadlikkuses ja turvakäitumises.

RIA 2024. aasta küberturvalisuse aastaraamatus [1] tuuakse välja, et kasutajate teadlikkus paroolide osas on viimase 5 aasta jooksul tõusnud (joonis 2). Probleeme valmistab parooli segmentide taaskasutamine. Näitena toodi välja ühe ettevõtte vastu teostatud õngitsusrünne, kus paluti sisestada oma praegune ning uus parool. Väideti, et esimene on lekkinud. Sealjuures salvestati algus- ning lõpusümbolid. Tuli välja, et 15 inimest pani oma uueks salasõnaks vana põhjal loodud parooli, vahetades ainult viimase sümboli, nt „Tallinn1" asemel „Tallinn2". Eelnev näide võiks anda piisava põhjenduse, et süsteemides tuleks suuremat tähelepanu pöörata korduvate mustrite kasutamisele, sest need võivad haavatavusi põhjustada.



Joonis 2. Küsimus: mida olete internetis või äpis isiklikul eesmärgil teinud turvalisuse või privaatsuse tagamiseks? [1]

Kristjan Pühveli bakalaureusetöös loodud ründesõnastik ülikooli parooliräside murdmiseks on konfidentsiaalne. Internetiotsingute tulemusena (kasutades otsingusõnu *paroolide sõnastik, eestikeelsed paroolid, estonian password list*) ei õnnestunud leida ka ühtegi avalikku viidet eestikeelsete ründesõnastike loomise ja kasutamise kohta.

4.2 Küsimustiku loomise alused

Praeguste hoiakute kaardistamiseks ning harjumuste väljaselgitamiseks koostati küsimustik, milles selgitati välja levinumad mustrid. Kaardistati Interneti-kasutajate paroolide kasutamise praktikaid eesmärgiga mõista inimeste arusaama tugevatest paroolidest. Saadud info põhjal antakse analüüsi peatükis ülevaade levinumatest mustritest, mis võivad turvaprobleeme põhjustada. Samuti täiendatakse eestikeelsete paroolide sõnastikku, et seda oleks võimalik kontot nõudvate keskkondade turvalisuse suurendamiseks kasutada. Küsimustiku üldisem eesmärk oli tõsta inimeste turvateadlikkust paroolide osas ning parandada küberhügieeni.

Küsitlusuuringu koostamiseks ning selle põhimõtete järgimiseks on kasutatud Tartu Ülikooli sotsiaalse analüüsi meetodite ja metodoloogia õpibaasi[14], kus on antud juhised küsitlusuuringu

[14] Sotsiaalse Analüüsi Meetodite ja Metodoloogia õpibaas. https://samm.ut.ee

teostamiseks. Kuna selle küsitlusuuringu teema on delikaatne, siis on vastajale rõhutatud uuringu anonüümsust ning vastamise vabatahtlikkust. Küsitlusuuring viidi läbi LimeSurvey[15] keskkonnas.

Ankeedi koostamisel on lähtutud mudelist tabelis 1 [17], kus küsimused on jaotatud temaatilistesse plokkidesse lähtuvalt vastava ploki eesmärgist ja proovitud jaotada nii, et igas plokis oleks võrdne arv küsimusi. Kõige alguses on nõusolekuvorm, mille täitamata jätmise või eitava vastuse puhul suunatakse vastaja küsimustiku lõppu. Nõusoleku vormi pole tabelis välja toodud. Küsitlusuuring täisversiooni leiab lisast 1.

Tabel 1. Küsimustiku struktuur, kohandatud [17]

Eesmärk	Küsimustiku teemad	Uuringu küsimused
Millised on kasutajate harjumused ja praegused paroolide mustrid?	1. Harjumused	1. Temaatiline plokk
	2. Paroolide genereerimine	2. Temaatiline plokk
	3. Stsenaariumid	3. Temaatiline plokk
	4. Demograafiline teave	4. Temaatiline plokk

Kasutajate praeguseid praktikaid kaardistati esimese plokis. Selle eesmärk oli teada saada, kui relevantne on sõnastiku loomine ja mustrite uurimine, sest kaheastmelise autentimise ja genereeritud paroolide kasutamise puhul poleks see aktuaalne. Järgnevalt uuriti, mis teemad ja valdkonnad küsitletuid paroolide loomisel inspireerivad ning mis keeles seda tehakse. Hindamaks, kas keeleline ja kultuuriline taust on olulised ja millises osakaalus. Parooligeneraatori kasutamise puhul ei oleks eelmärgitu tähtis. Kolmas plokk kätkes endas mängulisi stsenaariumeid mõistmaks, kuidas mõjutavad etteantud olud paroolide valimist. Näiteks oli palutud luua parool konkreetse (nimelise) veebikeskkonna (pank PayPal, meilirakendus Gmail vms) jaoks, et mõista kas veebikeskkonna kaalukus mõjutab paroolide loomist. Samuti, kuidas toimitakse, kui paroolide loomisel on nõuded ette antud, nt kaheksamärgiline ja peab sisaldama suurtähte. Ehk siis, kas pigem täidetakse miinimumnõudeid või lisatakse ise veel tugevust ja keerukust juurde. Uuriti ka

[15] LimeSurvey. https://www.limesurvey.org

osalenud inimeste demograafilist tausta, et seeläbi mõista tulemusi ning võimaldada nende paremat rakendamist.

Ühelegi küsimusele vastamine polnud kohustuslik, mis tähendab, et vastaja võis esitada ka pooleldi täidetud ankeedi. See mõjutas küll tulemuste analüüsi, kuid eesmärk oli koguda võimalikult palju erinevaid vastuseid (paroolide mustreid) ning ka üksikud vastused olid praegusel juhul tulemuste osas kasulikud.

Valim koosnes Tartu Ülikooli erinevate valdkondade (sotsiaalteadused; humanitaarteadused ja kunstid; meditsiiniteadused; loodus ja täppisteadused) esindajatest. Selle küsimustiku puhul oli mittetõenäosuslik valim kombinatsioon [18]:

- mugavusvalimist, mille moodustas koostajale lähedane üliõpilaskond;
- sihipärasest valimist, eesmärk oli, et esindatud oleksid kõik valdkonnad;
- lumepallivalim, sest paluti küsimustikku oma ringkonnas edasi jagada.

Kuna tegu oli mittetõenäosusliku ning mitteesindusliku valmiga, siis ei saa küsitlusuuringu tulemusi laiendada suuremale inimgrupile. Samas aitab see avada mõningaid kasutajate loodud paroolide mustreid, et neid sõnastiku genereerimisel kasutada.

Kuna paroolide lekked põhjustavad suurt majanduslikku kahju, siis küsitlusuuringust kogutud andmed võiksid aidata seda olukorda leevendada. Küsitlusuuringu vastuseid analüüsitakse vastavas peatükis.

4.3 Andmete kogumine Internetist

Lisaks paroolidele lekib igal aastal mitmeid delikaatseid isikuandmeid. Eestis võib hiljutise näitena võib tuua Apotheka ostu- ja isikuandmete lekkimise [19]. Kuigi tänapäeval hoiustatakse kasutajakontode andmeid enamasti räsitud kujul, siis esineb ikkagi lekkeid, kus on avalikuks tulnud lihttekstis kirjed. Näiteks eelmise aasta esimeses pooles lekkisid Eesti Ettevõtluskõrgkkool Mainori andmed, mis sisaldavad ca 15 000 vilistlase isiklikku infot (sh salasõna, aadressi, isikukoodi, telefoninumbrit). See info on siiani Internetis müügis [20], hinnaga 8 *#CR*'i (veebikeskkonna enda valuuta) või 260 dollarit (tehes kasutajakonto), et pääseda infole ligi. Autor otsustas need andmed soetamata jätta, sest see oleks otsene rahaline toetus kuritegevusele, samal põhjusel pole kasutatud tasulisi pimeveebi andmeid. Lekke näidisandmetes ei ole paroolid eriti

huvipakkuvad (nt *i8agvfj*, *1i1nsgk6*, *k3u04vr5*). Siiski on ilmne, et kasutaja võib endale määrata küll tugeva parooli, ent veebikeskkonna nõrkuse tõttu satub tema isiklik info ikkagi Internetti.

Pimeveeb moodustab ca 0.01% internetist, autentimist vajav süvaveeb ca 90% ning pinnaveeb ülejäänu [21]. Selles töös on põhiosa teabest kogutud keskkonna GitHub pinnaveebi repositooriumist[16]. Minimaalsel hulgal andmeid (alla 1%) tuli LeakOsint *bot*'ist (*programm, mis inimtoiminguid automatiseerib ja/või jäljendab*[17]), mis tänaseks on suhtluskeskkonnast Telegram[18] eemaldatud.

Kasutaja Hacxx Underground repositooriumis „files" [22] leidub rohkelt viiteid erinevatele pinnaveebis allalaetavatele andmetõmmistele. Mitmed andmebaasid on esitatud *combo list*'idena, kus salasõna esineb räsimata kujul. Sageli on lekke juurde märgitud ka aastaarv, kuid on põhjust arvata, et andmed on lekkinud varem ning need on mainitud aastal ühte faili kogutud.

Selles repositooriumis leidub sadu linke failidele, millest mitmed on aegunud. Ebamõistlik oleks olnud kõik andmebaasid läbi käia. Seega sai valiku kriteeriumiks andmelekke kirjeldus, mille alusel tuli järeldada, kas seal võib .ee domeeniga tunnuseid leiduda. Failide kirjelduste otsing andis tulemuseks vaid eek.ee (Mainor) keskkonna lekke. Kuigi läbitud sai rohkem faile, on välja toodud need andmelekked, kust leiti .ee domeeniga paroole (sulgudes on leitud paroolide arv) [22]:

1. *EMAIL:PASS - 1.4M HQ Combo List* (37)
2. *100k Mailaccess JUNE 2023_20* (20)
3. *600k Mail Access Combo Good For Everything DATE: 27.06.2020* (4235)
4. *838K [Fortnite,Spotify,NetFlix,Brazzers,Steam,PSN,Origin,Minecraft,and more....]* (1)
5. *Email:Pass 396k Combo Gaming RELEASE 29.06.2020* (201)
6. *10 Mil Fresh HQ SQL E-Pass Combo! - Released June 2020* (884)

Põhiosa paroolidest tuli 2017. aasta massiivsest andmelekkest, kus avalikustus lihttekstina 1,4 miljardi kasutaja info, RIA andmetel ca 198 000 .ee domeeniga kasutaja oma [23]. Käesoleva töö raames tuvastati lekkest 184 019 unikaalset parooli. Andmete eraldamiseks kasutatud koodi leiab lisast 1.

[16] GitHub. https://github.com
[17] Andmekaitse ja infoturbe portaal. Cybernetica AS. https://akit.cyber.ee
[18] Telegram. https://web.telegram.org

Huvialused objektid olid paroolid ja nende mustrid, seega kasutajanimesid ei kogutud ega salvestatud. Samuti oli eesmärk uurida unikaalseid paroole, mis tähendab, et kordusi ei salvestatud. Lõplik unikaalsus andmelekete vahel saavutati pärast nende omavahelist võrdlemist, mille tulemusena jäi alles 187 223 kirjet. Need annavad ülevaate salasõnade koostamise praktikatest minevikus ega kajasta hetkeolukorda.

4.4 Andmete kogumine veebilehelt

Lisaks leketest ning küsitlusuuringust saadud andmetele, koguti anonüümset kasutajasisendit ka loodud veebirakenduses Paroolikratt[19], et võrrelda seda aegunud andmete (lekete näol) ja kasutajate arvamusega enda paroolidest (küsitlusuuringu vastused). Veebilehel on lisatud märge mitte sisestada aktiivselt kasutuses olevaid salasõnu, mis tähendab, et sisend ei pruugi kajastada reaalselt kasutusesolevaid paroole. Veebilehel antakse ka nõuandeid tugeva parooli koostamiseks. Need on suures osas koostatud küberturvalisuse aastaraamatus 2024 [1] antud soovituste põhjal. Samuti on lisatud parooligeneraator, millega saab kasutaja 1-10- sõnalisi salafraase koostada. Sõnad valitakse juhuslikult ca 70 eestikeelse lemma hulgast.

Kuna sisendi andmine toimub mitteametlikus keskkonnas ning pole juurdeviivaid küsimusi, siis on põhjust arvata, et kasutajad sisestavad sinna testimiseks aegunud, laialt levinud või neid huvitava mustriga paroole. Võib arvata, et seda kasutatakse ka n-ö mängimiseks, et lekkinud paroole tuvastada. Umbes nädalal aja jooksul koguti 368 parooli, mille mustrite ja pikkuse analüüs toimub vastavas peatükis.

Sellega on antud ülevaade töös kasutatud andmete kogumise metoodikast. Infot koguti kolmel viisil – küsitlusuuring, paroolilekked ja veebilehe kasutajasisend. Eesmärk oli saada ülevaade aegunud mustritega paroolidest (lekked), inimeste hinnangust oma paroolidele (küsitlus) koos võimaliku lisasisendiga (veebirakendus). Nende andmete analüüsi ja tulemusi käsitletakse kuuendas ja seitsmendas peatükis.

[19] Paroolikratt. https://paroolikratt.ee

5. Veebilehe arendus

Töö käigus valmis veebileht, mille eesmärk oli testida kasutajasisendit kahe andmebaasi (lekkinud ning genereeritud paroolide ehk sõnastiku vastu). Samuti on välja toodud nõuanded paroolide loomiseks ning kasutajal on võimalik parooligeneraatoriga luua eestikeelne salafraasi. Veebilehe arendamisel on arvestatud funktsionaalsete ja mittefunktsionaalsete nõuetega. Järgnevalt tutvustatakse veebirakenduse loomise nõudeid, tehnoloogiad ja arhitektuuri.

5.1 Funktsionaalsed ja mittefunktsionaalsed nõuded

Konkreetseid nõudmised veebirakenduse Paroolikratt[20] disaini ja tehnoloogia osas polnud ette kirjutatud ja need tuli autoril endal defineerida. Veebilehe arendamisel järgiti sarnaste keskkondade stiili, nt Security.org[21] ja Have I been pwned?[22] Salafraasi generaatori puhul võeti eeskujuks Bitwardeni[23] ja rabool.eu[24] stiil. Lisaväärtusena kuvatakse kasutajale mõlema vahelehe alumises osas nõuandeid paroolide loomiseks ja hoiustamiseks.

Loodud veebilehel on inimestel võimalik paroole testida. See on tehtud eesmärgiga kasvatada küberteadlikkust ning luua diskussioon sarnaste veebikeskkondade eetilisuse ning turvalisuse osas. Samuti peegeldavad sisestatud paroolid inimeste harjumusi ja inspiratsiooniallikaid salasõnade loomisel. See võiks anda ettekujutuse ka levinud mustritest, mis võimaldab kaasajastada paroolide loomise soovitusi.

5.1.1 Funktsionaalsed nõuded

Funktsionaalsed nõuded on põhiliselt kasutajapoolsed nõuded, mis kirjeldavad veebilehe funktsioone ehk mida see peaks tegema [24] . Siinpuhul olid olulised järgmised aspektid:

1. Pealeht.
    1. Sisaldab ala parooli sisestamiseks, mispeale kuvatakse vastus (lekkinud, genereeritud või ei leitud).
    2. Kuvatakse soovitusi paroolide loomise osas.

[20] Paroolikratt. https://paroolikratt.ee
[21] Security.org. https://www.security.org/how-secure-is-my-password
[22] Have I been pwned? https://haveibeenpwned.com
[23] Bitwarden. https://bitwarden.com/password-generator/
[24] rabool.eu. https://rabool.eu/

2. Generaatori vaheleht.
    1. Ala, kus kuvatakse loodud salasõna.
    2. Kasutaja saaks muuta fraasi pikkus, lisada suurtähti ja numbreid.
    3. Võimalus salafraas kopeerida.
    4. Kuvatakse soovitusi paroolide loomise osas.

Eelnimetatud nõudeid rakendati enne mittefunktsionaalseid nõudeid.

5.1.2 Mittefunktsionaalsed nõuded

Mittefunktsionaalsed ehk kvaliteedinõuded määravad süsteemi omadused ehk kuidas veebileht töötab ja vastavus kasutaja ootustele [24].

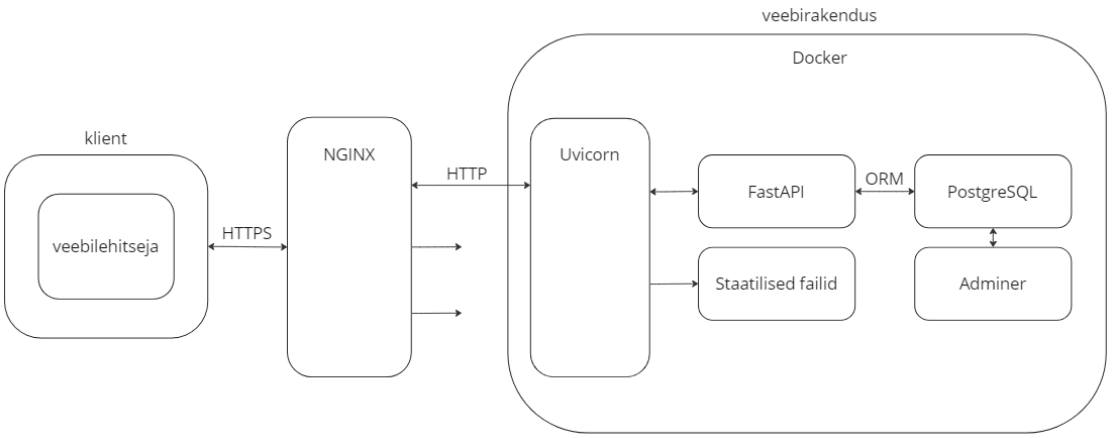
1. Andmebaasipäring peab olema võimalikult kiire.
2. Päring ja vastused peavad olema krüpteeritud, et vältida ründeid.
3. Veebileht peab olema kasutatav nii mobiilis kui ka *desktop* arvutis.
4. Veebileht peab olema kasutatav erinevates veebibrauserites.
5. Veebirakenduse nimi peab olema lihtne, kuid lööv.
6. Veebilehe peamised tekstid peavad olema loetavad ehk kontrasti vastavus standardile[25].
7. Veebilehe tagaliidest peab olema kerge hallata.
8. Veebileht peab vastama veebilehitseja jõudlustestidele vähemalt 90%.
9. Veebileht peab olema esteetiline (värvivalik vastaks temaatikale).

Nõuded on esitatud olulisuse kasvamise järjekorras.

5.2 Arhitektuur

Veebirakenduse Paroolikratt arhitektuuri disainimisel on lähtutud RESTful (*Representational State Transfer*) stiilist ehk server teenindab iga kliendi päringuid eraldiseisvalt [25]. Rakendus ise koosneb kolmest peamisest Dockeri konteinerites töötavatest komponendist: Uvicorn[26] serverist koos FastAPI[27] rakendusliidesega, PostgreSQL[28] andmebaasist ja Adminer[29] andmebaasihaldurist.

[25] Web Content Accessibility Guidelines (WCAG) 2.2. https://www.w3.org/TR/WCAG22
[26] Uvicorn. https://www.uvicorn.org
[27] FastAPI. https://fastapi.tiangolo.com
[28] PostgreSQL. https://www.postgresql.org
[29] Adminer. https://www.adminer.org



Joonis 3. Veebirakenduse arhitektuur.

Veebirakenduse Paroolikratt arhitektuuri on kujutatud joonisel 3, kus üks osapool on klient, kes kasutab veebilehitsejat, mis HTTPS (*Hypertext Transfer Protocol Secure*) päringu kaudu suhtleb NGINX[30] pöördproksiga, mis omakorda suhtleb Uvicorn serveriga protokolli HTTP (*Hypertext Transfer Protocol*) kasutades. Uvicorn edastab veebilehitsejale sisu kuvamiseks vajalikke staatilisi faile (CSS, HTML, JavaScript). Uvicorn ja FastAPi suhtlevad omavahel sisemisi ühendusi kasutades. Andmebaasi info liigub SQLAlchemy[31] moodustatud päringutega Psycopg 2[32] toel andmebaasi ja rakendusliidese vahel. PostgreSQL andmebaasi sisu kuvatakse haldajale interaktiivselt Admineri toel.

5.3 Tehnoloogiad

Veebilehte arendades jäid kavandatud tehnoloogiad peamiselt samaks, sest suuremaid tagasilööke ei esinenud. Tehnoloogiate osas olid mõningad alternatiivid, kuid tehtud valikuid on järgnevates alapeatükkides ka põhjendatud. Veebirakendus sai suures osas üles seatud järgides Kevin Kimi [26] blogipostitust *A Guide to Connecting PostgreSQL and Pythons' Fast API: From Installation to Integration* Medium veebikeskkonnas. Loodud veebirakenduse koodi leiab lisast 1.

5.3.1 Konteineritehnoloogia

Docker[33] on avatud platvorm, mis lubab eraldada rakenduse riistvarast ning muudab seeläbi implementatsiooni lihtsamaks. Dockeri konteineri sees on Docker Image, mis luuakse Dockeri

[30] NGINX. https://www.nginx.com
[31] SQLAlchemy. https://www.sqlalchemy.org
[32] Psycopg. https://www.psycopg.org
[33] Docker. https://www.docker.com

faili põhjal. Veebirakendus Paroolikratt on ehitatud Dockerit kasutades, see koosneb kolmest konteinerist (server, andmebaas ning andmebaasi tööriist). Praegusel juhul on Dockeri eelised kergekaalulisus, seadistamise, muutmise ja liigutamise lihtsus ning isoleeritus ehk kapseldamine ja turvalisus.

5.3.2 Veebirakendus

Konkreetselt eristatavat eesliidest ja tagaliidest FastAPI' kasutamise puhul välja tuua ei saa, mida on näha ka jooniselt 3. Rakenduse dünaamilise sisu loomiseks kasutatud Vanilla JavaScripti. Veebilehe märgistuskeelena kasutati HTML5 (*HyperText Markup Language*) ja visuaali kujundamiseks märgendkeelt CSS (*Cascading Style Sheets*). Paroolikrati veebirakendus on oma olemuselt lihtsa struktuuri ning funktsionaalsustega, mistõttu ei olnud vajalik ühegi raamistiku (nt Angular[34]) rakendamine. Päringute haldamisega seotud pool koosneb kolmest peamisest osast – veebiserver, API (*Application Programming Interface*) ehk rakendusliides ja andmebaas.

Uvicorn on ASGI (*Asynchronous Server Gateway Interface*) veebiserver programmeerimiskeelele Python ehk liides, mis vahendab veebibrauseri ühendust ja laseb rakendusliidesel FastAPI päringuid käidelda. Ka FastAPI ise soovitab rakenduse jooksutamiseks Uvicorni [27]. Uvicorn teenindab selles rakenduses nii tagaliidest kui ka staatiliste failidega eesliidest.

Veebirakenduse loomiseks on kasutatud rakendusliidest FastAPI [28], sest see on lihtsasti rakendatav (lihtne loogika ja tuvastab vigu), kergesti omandatav (Python'il põhinev), vabavaraline ja turvaline. Valikus oleks olnud ka Flask[35], kuid FastAPI pakub:

- *full-stack* raamistikku;
- asünkroonseid rakendusi (Flask on WSGI (*Web Server Gateway Interface*));
- paremat kiirust;
- sisseehitatud andmete valideerimist.

Miinuspoolena võib välja tuua, et kuna see on uuem, siis ei ole nii laiapinnalist tuge juhendite ja foorumite näol [29]. FastAPI' Python'i andmete valideerimiseks ja mudelite loomiseks on kasutatud Pydantic'ut[36].

[34] Angular. https://angular.io
[35] Flask. https://flask.palletsprojects.com/en/3.0.x
[36] Pydantic. https://docs.pydantic.dev/latest

5.3.3 Andmebaas

Andmebaasina oli alguses plaanis kasutada MongoDB , kuid see otsustati andmebaasi süsteemi PostgreSQL vastu välja vahetada, sest viimast oli hõlpsam rakendada. Samuti ei olnud kasutatavate andmete puhul andmebaasiliideste funktsionaalsuses erinevusi, st polnud oluline, et andmebaas oleks mitterelatsiooniline.

PostgreSQL[37] on avatud lähtekodiga relatsiooniline andmebaasi haldamise süsteem, mille eelised on rakendamise lihtsus ja vabavaralisus. Viimast mainitakse ka puudusena, lisaks veel väiksemat kiirust kui andmebaasi süsteemil MySQL[38] [30]. Kuid praegusel juhul kasutatavad relatsioonid on väga algelised ning tabelid lihtsakoelised, sisaldades peamiselt vaid paroolikirjet ning ID'd (primaarvõti). Seega pole põhjust arvata, et siinpuhul oleks teise tarkvara kasutamine olulise eelise andnud. Kõik andmebaasi paroolikirjed on lisatud BASE64 kodeeringus vältimaks sõnastiku lekkimist lihttekstina. Tehtud valik suurendab küll andmebaasi mahtu ning vähendab otsingukiirust, kuid lisab turvalisust. Andmebaasile pääseb ligi ainult sisevõrgust.

Andmebaasi kasutuse lihtsustamiseks, Python objektid tõlgitakse andmebaasi andmetüüpideks, on kasutatud teeki SQLAlchemy[39], mis on ORM (*Object Relational Mapper*) instrumentaarium (*toolkit*) programeerimiskeelele Python genereerides SQL päringuid. Andmebaasi adapterina on kasutusel Psycopg 2[40], mis suhtleb andmebaasiga ehk edastab päringuid.

Haldamiseks ning andmebaasi sisu lihtsamaks kuvamiseks on rakendatud skriptimiskeelt PHP (*Hypertext Preprocessor*) kirjutatud andmebaasi halduse tööriista Adminer[41], mis käivitatakse eraldi Dockeri konteineris.

Andmebaasi on lisatud ca 50 miljonit kirjet, millest 49 927 477 on genereeritud paroolid, 187 223 lekkinud salasõnad ning ca 400 kirjet kasutajasisend. Andmebaasi lisatud genereeritud paroolide failide maht on ca 1 GB, mis on eeldatust 20 korda vähem (kokku sai genereeritud üle 53 GB). Otsingukiiruse säilitamise huvides sai lisatud vaid järgnevaid mustrid: 1) Lemma[1-100] 2) Niminimi! 3) nimi[1-999] 4) nimi[1900-2024]. 5) nimi[1900-2024]! 6) lemma Praegusel juhul võib otsing kohati 3-4 sekundit aega võtta.

[37] PostgreSQL. https://www.postgresql.org
[38] MySQL. https://www.mysql.com
[39] SQLAlchemy. https://www.sqlalchemy.org
[40] Psycopg. https://www.psycopg.org
[41] Adminer. https://www.adminer.org

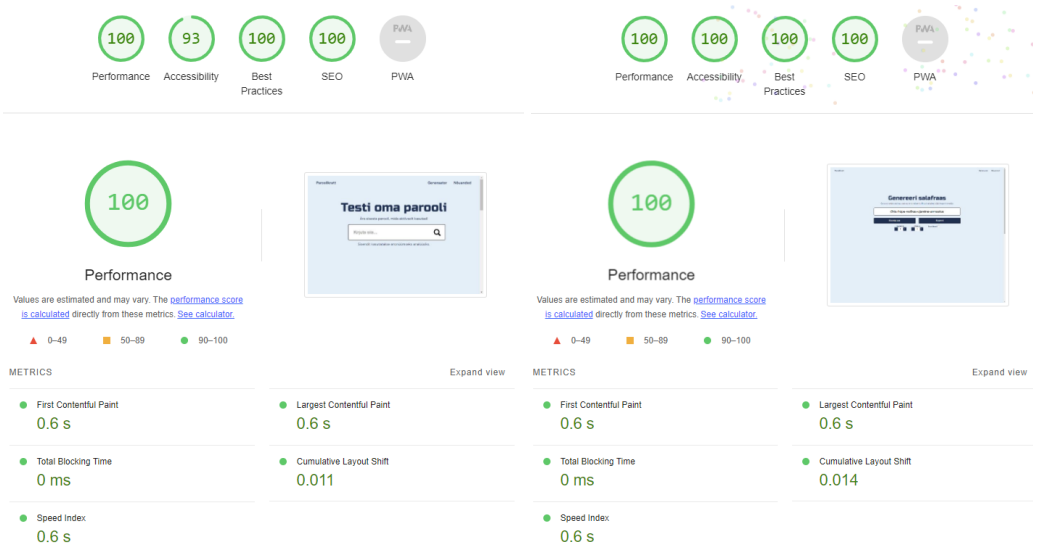
5.3.4 Domeeni registreerimine ja serveri seadistamine

Veebilehe avalikustamiseks on kasutatud domeeni paroolikratt.ee. Nimi sai inspiratsiooni eesti folkloorist, kus kratt kujutab endast *nõiduslikku olendit, kes on valmistatud vanadest esemetest vm ning kes tule- või sädemejoana ringi lennates vara kokku veab.*[42] Enne registreerimist kinnitati domeeni vakantsus Eesti Interneti SA lehel[43] ja seejärel registreeriti domeen Zone[44] keskkonnas.

Veebirakendust hoiustatakse Richard Jõgi isiklikus Ubuntu 22.04[45] operatsioonisüsteemi kasutavas serveris. R. Jõgi andis nõu rakenduse portide vahetamise osas, et need ei kattuks tema serveris olevate teiste programmidega. Tarkvaras NGINX Proxy Manager[46] seadistatud vajalikud turvakonfiguratsioonid, nt SSL (*Secure Sockets Layer*).

5.4 Testid

Nii pealehe kui ka generaatori vahelehe testimiseks on kasutatud veebilehitsejasse Chrome sisseehitatud avatud lähtekoodiga laiendust Lighthouse[47], mis pakub arendajale tuge rakenduse parandamiseks. See analüüsib nelja peamist aspekti (jõudlus *perfomance*, ligipääsetavus *accessibility*, head tavad *Best Practices* ja nähtavust otsingutes *SEO*).



Joonis 4. Paroolikratt.ee Lighthouse testide tulemused (pealeht ja generaator vaheleht).

[42] Sõnaveeb. https://sonaveeb.ee
[43] Eesti Interneti SA. https://www.internet.ee
[44] Zone. https://www.zone.ee
[45] Get Ubuntu Server. Ubuntu. https://ubuntu.com/download/server
[46] Nginx Proxy Manager. https://nginxproxymanager.com
[47] Chrome Lighthouse. https://developer.chrome.com/docs/lighthouse/overview

Jooniselt 4 on näha, et esilehe puhul said kõik testid peale *Accesibilty* maksimaalse tulemuse. Ligipääsetavuse skoori vähendamise põhjuseks olid *links do not have a discernible name*, mis ei oma veebilehe funktsionaalsuse osas suuremat mõju. Joonisel 4 on näha, et generaatori vahelehe kõikide testide tulemused olid maksimaalsed.

5.5 Disain

Veebilehel värvikontrast vastab WCAG 2.2 (*Web Content Accessibility Guidelines*) nõuetele. Põhiteksti ning taustavärvi kontrast on 11.19, mis vastab AAA tasemele.[48] Värvigamma on valitud sinistes toonides, sest sinine sisendab usaldusväärsust ning rahulikkust [31]. Värvipaletilt on valitud üksteisega analoogsed värvid kasutades väikeste muudatusega Color Designeri eelgenereeritud värviskeemi[49], mida on kujutatud joonisel 5.

Joonis 5. Värviskeem (Color Designer; kohandatud)

Kontrasti lisab lekkinud või genereeritud parooli sisestamisel kuvatav oranž, mis on paletilt valitud sinise vastandvärvina ja näitab ohtu. Rohelist kuvatakse, kui parooli ei leitud ühestki andmebaasist, see on turvatunde indikatsioon.

5.6 Salafraaside generaator

Eestikeelne salafraaside generaator on loodud Bitwarden'i[50] eeskujul. Kasutajal on võimalik genereerida 1-10 sõnaline salafraas, kus sõnade eraldajaks on sidekriips. Keerukuse kasvatamiseks on võimalik valida läbiv suurtäht. Tugevuse suurendamiseks on võimalik lisada sõnade lõppu juhuslikud numbrid. Sõnades ei tehta numbritega täheasendusi, vaid lisatakse need sõnade lõppu, sest eesmärk on tugevuse mitte keerukuse kasv. Ühesõnalise parooli genereerimine on kasutaja enda vastutusel, veebisaidil on välja toodud soovitused paroolide koostamiseks ja hoiustamiseks.

[48] Colour Contrast Checker. https://colourcontrast.cc
[49] Color Designer. https://colordesigner.io/color-scheme-builder#E1E6FA-C4D7ED-ABC8E2-375D81-183152
[50] Strong password generaator. https://bitwarden.com/password-generator

6. Analüüs

Töö käigus koguti andmeid kolmel viisil (lekked, küsitlusuuring ja kasutajasisend). Järgnevalt uuritakse kõiki neid aspekte nii eraldi kui ka tehakse üldine analüüs. Selle jaoks on peamiselt kasutatud Pythonit, selle andmetöötlusteeki Pandas[51] ja tarkavara Jupyter Notebook[52].

6.1 Küsitlusuuring

Küsitlusuuring oli vastamiseks avatud ca 1 kuu, selle aja jooksul laekus 225 vastust, millest 202 lubati bakalaureusetöö jaoks kasutada. Igale küsimusele vastamine oli vabatahtlik, mis tähendab, et vastaja võis esitada ka poolikult täidetud ankeedi. Vastamisega jõudsid lõpuni 109 inimest 202st. Küsimustiku peamine eesmärk oli saada infot paroolide mustrite kohta (andmete kogumine), mitte teostada laiaulatuslik sotsiaaluuringu printsiipidest lähtuv analüüs.

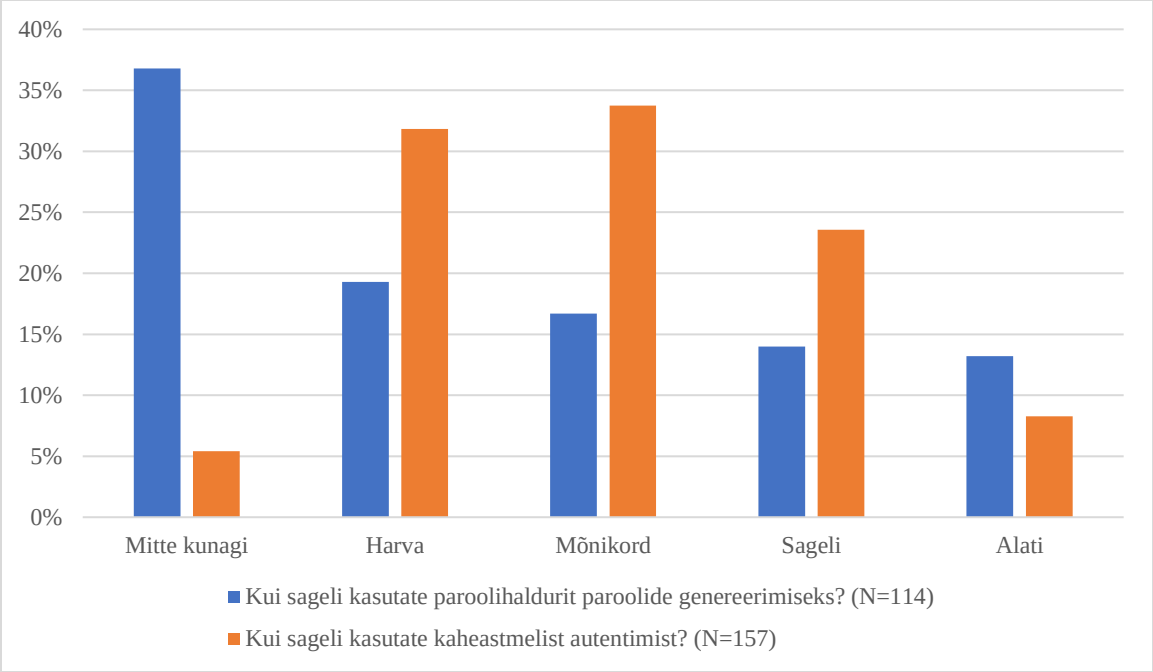
Järgnevalt on igat teemaplokki käsitletud eraldi alapeatükis, analüüsi juurde on lisatud joonised. Välja on toodud huvitavamate vastuste diagrammid. Vastamatajätmistega pole enamasti diagrammil kujutamisel arvestatud, sulgudes on välja toodud vastuste hulk ning y-teljel on enamasti kujutatud osakaal protsentides. Kui pole öeldud teisiti.

6.1.1 Harjumused

Vastanud hindavad oma oskust Internetis tegutseda pigem heaks. Samas tasub arvestada, et valikuvariante oli viis, esines keskmise valik „hea“ (saab tõlgendada neutraalsena), mille oli märkinud 45% (72). Kaheastmelise autentimise puhul on märgata, et seda kasutatakse, aga harjumus pole sügavalt juurdunud (joonis 6). Sageli või alati kasutab seda ca 31,8% (50) vastanutest, mõnikord ca 33,8% (53) ja ülejäänud ei kasuta üldse või harva. See tähendab, et potentsiaalselt võivad nende paroolid mitmes keskkonnas haavatavad olla.

Vastanutest 72,6% (114) kasutavad paroolihaldurit. Ühelt poolt võib see viidata, et kasutusel rohkelt (erinevaid) salasõnu, mida ei ole võimalik meeles hoida. Teisalt pole kindel, mil viisil need halduris hoiustatud on, ka seal võib olla lekkeid. Samuti ei pruugi nad pöörata tähelepanu halduripoolsetele soovitustele, nt kui parool leitakse lekkinud andmete hulgast või paroolid on liiga sarnased. Samas salasõna genereerimiseks paroolihaldurit pigem ei kasutata, mis viitab sellele, et paroolid luuakse omaloodud mustrite ja seoste põhjal. Tulemused on näha joonisel 6.

[51] Pandas. https://pandas.pydata.org
[52] Jupyter. https://jupyter.org

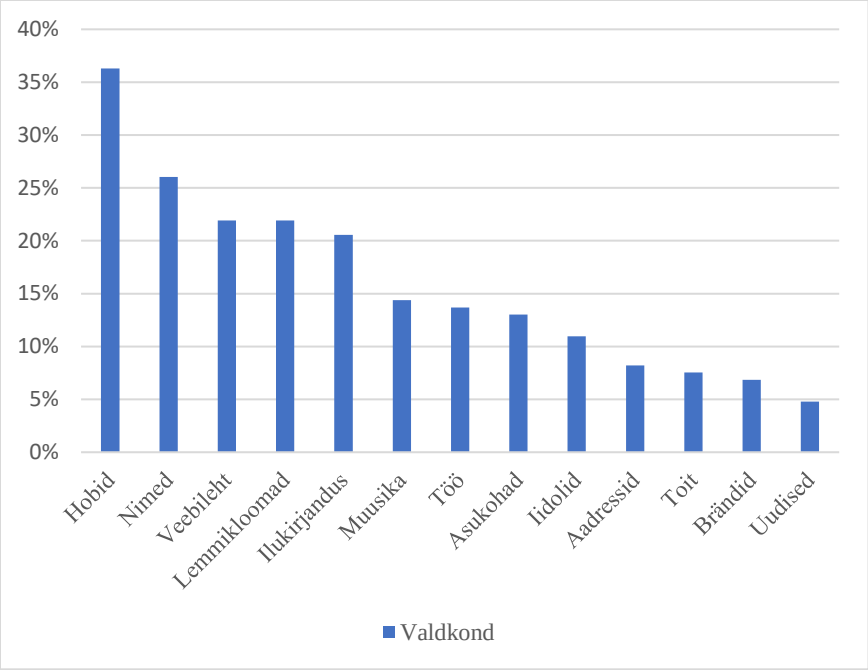


Joonis 6. Parooligeneraatori ja kaheastmelise autentimise kasutamine.

Kõige turvalisemaks hoiustamise viisiks peetakse igal sisselogimisel uue salasõna genereerimist 44,7% (68) ehk n-ö ühekordseid salasõnasid või enda meeles hoitud paroole 41,4% (63). Kõige vähematurvalisemaks peeti meilile saadetud 45,3% (68) või pilves hoiustatavasse dokumenti salvestatud paroole 23,3% (35).

6.1.2 Paroolide genereerimine

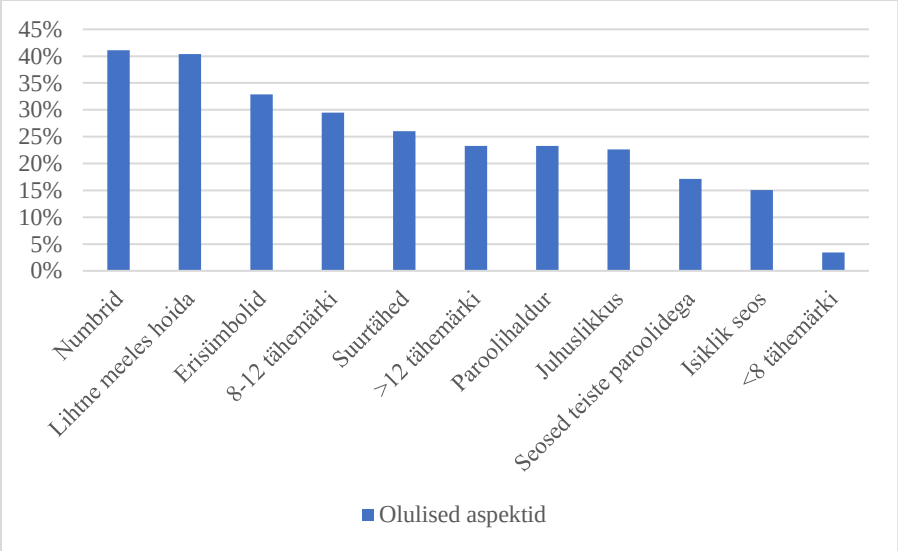
Jooniselt 7 on näha, et kõige enam inspireerivad vastanuid salasõnade loomisel hobid, seejärel nimed ja veebilehed, kuhu konto tehakse. Samuti toodi välja, et mitmel juhul tuleb inspiratsioon laual olevatest esemetest või juhuslikest asjadest, roppustest, nimedest ja kuupäevadest. Sõnastiku koostamisel annab see olulist infot, milliseid teemasid katta.



Joonis 7. Inspiratsiooniallikad paroolide loomisel (N=146).

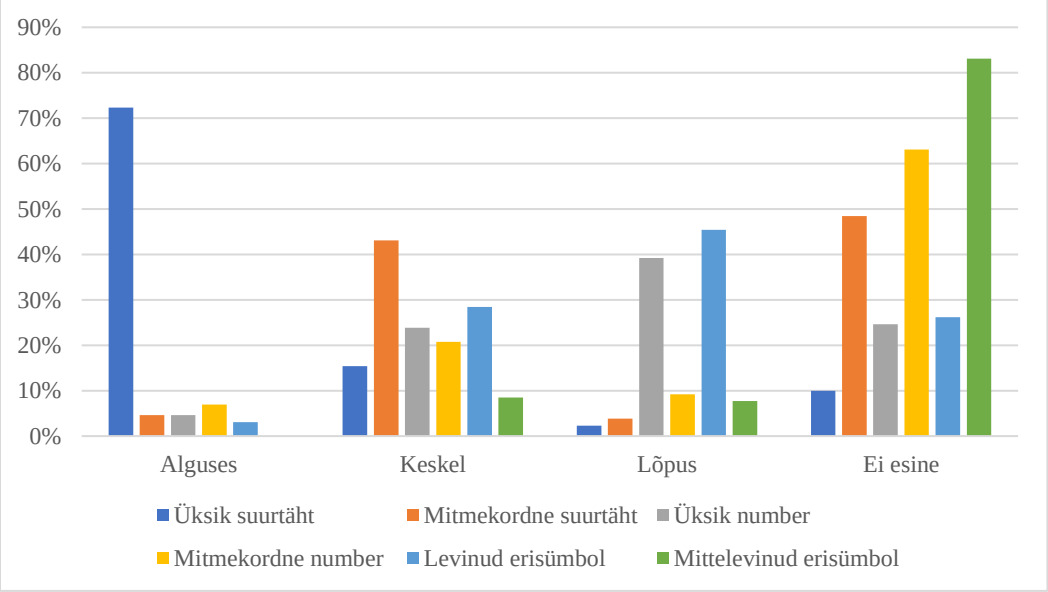
Paroolide loomisel peetakse oluliseks üldlevinud soovitusi [32], mis vastavad keskkondade salasõnade nõudmistele. Ehk siis parooli peab olema lihtne meeles hoida, see peab sisaldama erisümboleid, numbreid ja olema 8-12 tähemärki pikk, ka suurtähtede sisaldumine märgiti ära. Vastused on toodud joonisel 8. Isiklikku seost või seost teiste paroolidega ei peeta oluliseks, samuti ei ole paroolid alla 8-märglised. Pikkust ja seos teiste paroolidega on süsteemil või paroolihalduril võimalik kontrollida, aga isiklikku seost mitte.

Stsenaariumitega küsimuste puhul sisestati põhiliselt 8-16 märgilisi paroolinäiteid (joonis 10). Kõige populaarsemad esimese ja viimase sümboli kombinatsioonid (joonis 11) on sellised, kus suurtäht on alguses ja number, sümbol või väiketäht lõpus, mis kinnitab paroolide loomisel oluliseks peetud aspekte (joonis 8). Ka esitähe ja kolme viimase sümboli analüüsis on kõige sagedam ainult numbritest, väiketähtedest või kahest numbrist ja sümbolist koosnev lõpp. Esimene märk on enamasti suurtäht.



Joonis 8. Mida peate oma paroolide koostamisel oluliseks? (N=146).

Enda sõnul vastajad pigem ei kasuta eestikeelsed või sellest tuletatud paroole. Vaid 9,9% (14) oli märkinud, et nad teevad seda alati, kuid 33,8% (48) mitte kunagi. Samale küsimusele inglise keele kohta vastati, et alati kasutavad seda 2,9% (4) ja mitte kunagi 48,6% (68). Saadud vastused viitavad, et parool võidakse genereerida (ehk lihtsalt märgijada) või valitakse selliseid ühendeid, mida ei osata keelega seostada, nt keskkondade nimed või fiktsionaalsed tegelased või maailmad. Paroolinäidete põhjal oleks keelelist kuuluvust saanud kindlaks teha masinõppega või neid ükshaaval liigitades, mis oli praegusest uurimisteemast väljas.



Joonis 9. Valige, mis kirjeldab kõige paremini Teie põhiparoole (N=130).

Kui vastanutel paluti iseloomustada oma põhiparoole, siis selgus, et ka need vastavad üldjoontes peamistele soovitustele [32] ja nõuetele. Tulemusi on näha joonisel 9. Mustreid iseloomustab see, et üksik suurtäht paikneb enamasti alguses 72,3% (94), mitmekordne number lõpus 63,9% (82). Samas, kui mitmekordne suurtäht esineb, siis keskel 43,1% (56). See tõstab parooli keerukust. Samuti esines üksik number põhiliselt nii lõpus 39,2% (51) kui ka keskel, kus on sellega sõnastiku genereerimisel raskem arvestada. Levinud erisümboleid (. , ! ? # _ @) lisatakse nii lõppu 45,4% (59) kui ka keskele 28,5% (37), kus jällegi on neid keerulisem tuvastada. Mittelevinud erisümboleid ([ ] ;) pigem ei esine. See kajastus ka sisestatud paroolinäidetest, kus kõige levinumad lõpusümbolid olid hüüumärk ja punkt ning keskel esinesid enamasti mõttekriips, punkt ja alakriips. Loomulikult esineb ka kõikvõimalikke teisi erisümboleid, need lisavad paroolile keerukust. Sümbolite esinemine on välja toodud tabelis 2.

Tabel 2. Sümbolite esinemine lekkinud paroolide (N= 187 223), küsitluse sisendi (N=494) ja veebilehe sisendi (N=268) hulgas

Esimene sümbol Lekkinud/Küsitlus/Veebileht						Sümbolid sõna keskel Lekkinud/Küsitlus/Veebileht						Lõpusümbol Lekkinud/Küsitlus/Veebileht					
-	50	!	6	'	1	_	3600	-	61	tühik	11	!	416	!	69	!	7
(	48	?	4	$	1	.	1548	!	53	,	5	.	278	?	15	.	3
:	42	+	3			:	409	.	51	;	4	)	157	.	7	+	1
!	35	-	3			-	366	?	48	:	3	*	111	@	4	&	1
_	24	& [	2 2			&	352	_	37	=	3	:	106	#	4	@ $ …	1 1 1

Tabelis 2 on ära toodud iga andmestiku levinumate sümbolite esinemissagedused vastavalt parooli alguses, keskel või lõpus. Tulemuste tõlgendamisel tasub arvestada, et välja on toodud korduste arv ehk sõna keskel võis ka mitu sümbolit olla. Mittelevinud sümbolitest esines veel näiteks °, », ¤, £, >.

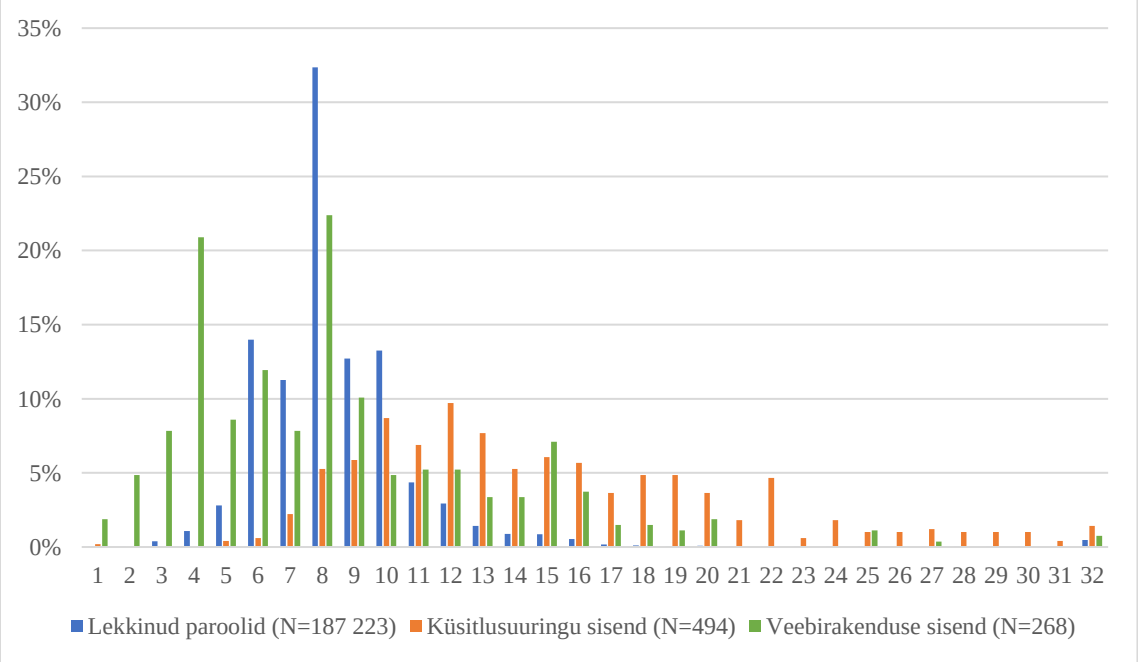
6.1.3 Stsenaariumid

Kui paluti mustri põhjal määrata, milliseid salasõnu kaalutaks enda parooliks, siis ilmnes, et valituks osutus peamiselt genereeritud parool *q+y3ncxy3w&}i+m* või salafraas *amiable-divinely-overrule-gossip-moustache*. Need valiks enda parooliks iga kord või sageli vastavalt 36,8% (42) ja 27,2% (31) vastanutest. Väljatoodud paroolid olid ka pikimad valikute hulgas. Tulemust võib põhjendada välistamismeetodiga, kui vastaja võrdles paroole omavahel, jättes kõrvale ilmselgelt nõrgad kirjeid, nt *qwerty*.

K. Pühveli uurimusest [6] tuli välja, et kolm kõige levinumat murtud paroolide mustrit olid kujul:

- Suurtäht, seitse väiketähte, kaks numbrit, 8,2% (121);
- Suurtäht, kaheksa väiketähte, kaks numbrit, 6,2% (91);
- Suurtäht, kuus väiketähte, kaks numbrit, 4,4% (65).

Küsitlusuuringu vastustest kindlaid mustreid välja ei joonistunud, st enamik paroole olid unikaalse mustriga. Selge erinevus ilmneb, kui võrrelda lekkinud paroolide pikkuseid küsimustiku sisendiga. Pikkused on toodud joonisel 10, kus on kuvatud kuni 32-märgiliste paroolide pikkuste esinemuse osakaal, sest sellest pikemate hulk oli marginaalne.



Joonis 10. Unikaalsete paroolide pikkus.

Paroole võrreldakse täiendavalt veel lekkinud paroolide ja veebilehe tulemuste alapeatükis.

6.1.4 Demograafiline teave

Vastajate sugu, emakeelt ja valdkonda uuriti viimases plokis. Vastustest ilmneb, et paljud vastajad nende küsimusteni ei jõudnud või ei soovinud vasta. Vastamata jäeti vastavalt 47,0% (95), 45,5% (92) ja 46,0% (93) . Sellest võib järeldada, et koostatud küsitlus oli liiga pikk ja vastajad kaotasid huvi. Samuti on oma osa sellel, et vastamine polnud kohustuslik. Selle ploki vastuseid tulemuste tõlgendamisel ei kasutatud.

6.2 Lekkinud paroolid

Erinevatest andmeleketest koguti kokku 187 223 eestikeelese keskkonna kasutajakontoga seostavat parooli, mis esinesid leketes tekstilisel kujul. Lekkinud paroolide andmestiku põhjal võib üsna kindlalt väita, et tegu oli aegunud nõuetega keskkonnast pärinenud mustritega ehk need paroolid on vananenud. Näiteks on lekkinud paroolidest kõige populaarsemad mustrid:

- Kaheksa väiketähte, 13,8% (25 926);
- Kuus väiketähte, 5,3% (9840);
- Seitse väiketähte, 3,8% (7049).

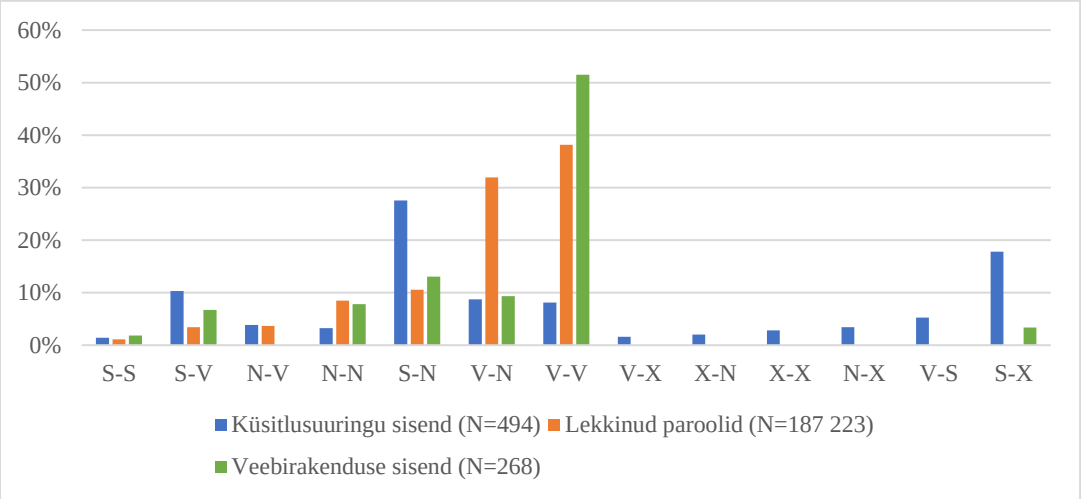
Praegusel ajal lasevad väga vähesed veebikeskkonnad sellise mustriga paroole määrata. Sama ilmneb paroolide pikkust analüüsides, kus põhiosa paroolidest jääb 6-10 märgi vahele ning 8-märglised moodustavad 32,4% lekkinud paroolidest (joonis 10).

Mustreid analüüsiti asendustega, kus on kasutatud regulaaravaldiste stiili:

- S – [A-Z];
- Ö – [ÄÕÖÜŽŠ];
- U – [A-ЯЁЃЂЊISЋЏЌЄЇ];
- V – [a-z];
- Ä – [äõöüžš];
- L – [a-яà-ÿёїґµћђјђґєњљіўsќџ];
- N – [0-9];
- X – [\W_ ³].

Jooniselt 11 on näha, et lekkinud paroolide puhul kõige populaarsemad mustrid algasid ning lõppesid väiketähega, samuti oli populaarne väiketäht alguses ja number lõpus kombinatsioon.

Suurtäht-number mustrid esines osakaaluna märkimisväärselt vähem. Kõikidest lekkinud salasõnadest oli esimese märgina kõige populaarsem väiketäht ning viimase märgina number. Viimase sümbolina esines kõige enam hüüumärk ning punkt (tabel 2). Ning parooli keskel olevatest sümbolitest olid populaarsed alakriips ning punkt.



Joonis 11. Unikaalsete paroolide esimene ja viimane sümbol.

Sümbolite valiku osas esineb mõningasi sarnasusi küsimustikku sisestatud näidetega (tabel 2), muus osas on kahe andmestiku tunnused võrdlemisi erinevad, mida võib põhjendada lekkinud paroolide vananenud andmestikuga. Samuti võib seda põhjendada küsitlusuuringu valimiga, kus olid enda hinnangult põhiliselt heade või väga heade Interneti-alaste teadmistega inimesed, kelle üldine küberteadlikkus on parem.

Veebirakenduse sisendi ja lekkinud paroolide puhul esines sarnasusi peamiste mustrite, pikkuse ja lõpusümbolite osas. Seda saab põhjendada veebisaidile sisendit andva kasutaja sooviga testida üldlevinud paroole (mineviku mustrite põhjal).

6.3 Veebilehe tulemused

Veebilehelt kogutud anonüümset sisendit hoitakse privaatsest andmebaasis krüpteeritud kuju. Uuriti, kas sisend kattub genereeritud või lekkinud paroolide andmebaasiga. Ca nädala jooksul antud 400 sisendist jäi pärast andmete esialgset puhastamist alles 368 kirjet, millest 268 kirjet olid unikaalsed. Joonistel 10 ja 11 ning tabelis 2 on andmed toodud unikaalsete kirjete kohta (N=268). Selle lõigu analüüs on kogu sisendi kohta (N=368). Emmast-kummast andmebaasist leiti vaste 45,9% (169). Selgus, et 19,0% (70) sisendist kattus genereeritud paroolide sõnastikuga, nendest

42,9% (30) olid unikaalsed. Võrdlemisi väikest katvust saab põhjendada võrdlemisi väiksest andmete hulgast, mis sai andmebaasi üles laetud (otsinguaeg oleks liiga pikk). Lekkinud paroolide hulgast leitud sõnesid oli 38,9% (143), nendest omakorda unikaalseid oli 57,3% (82). Kahe andmestiku peale kokku oli 92 unikaalset sisendit. Kummagi andmebaasiga mittekattuvaid unikaalseid sisendeid oli 65,7% (176).

Levinumad mustrid olid:

- Neli väiketähte, 6,7% (18);
- Kaheksa väiketähte, 6,3% (17);
- Viis või kuus väiketähte, 5,2% (14).

Nagu jooniselt 10 näha, siis sisendis domineerivad 4-9 märgilised kirjed. Jooniselt 11 tuleb välja, et kõige populaarsemad esimese ja viimase märgi mustrid olid väiketäht-väiketäht 51,5% (138), suurtäht-number 13,1% (35) ja väiketäht-number 9,3% (25). Mida võib seostada sellega, et veebirakendust kasutati põhiliselt n-ö testimiseks, et vaadata lihtsaid üldlevinud sõnu ja mustreid. Mis on positiivne indikaator, sest annab alust arvata, et inimesed ei usalda oma paroole suvalistele saitidele sisestada. Samuti võib oletada, et veebilehel olevad nõuanded ning generaator juhivad tähelepanu tugevale parooliloomise praktikale.

Analüüsi kokkuvõttena võib väita, kuigi lekkinud paroolidest ja kasutajasisendist ilmnes, et põhiliselt kasutatakse paroolides väiketähti, tulenes see peamiselt aegunud sisendist ja kasutajate usaldamatusest veebirakenduse osas. Arvestades küsitlusuuringu tulemusi koos veebikeskkondade parooliloomise nõuetega (E-ITS[53]) siis saab öelda, et vastanud lähtuvad paroolide loomisel sageli levinud mustritest. Mitmel juhul algab parool suurtähega, millele järgnevad väiketähed ning lõpus on numbrid ja/või erisümbolid. Inspiratsiooni saadakse nimedest (isiklikest või veebikeskkonna omast) ning suvalistest asjadest ehk lihtsalt nimisõnad. Keerukam on tuvastada neid mustreid, kus ei järgita üldist reeglipära, mis tähendab, et suurtähed lisatakse keskele või lõppu, sõnaosasid ühendatakse erisümbolitega ja numbreid kasutatakse ka keskel või alguses. Järgmises peatükis antakse ülevaade levinud mustrite alusel sõnastiku loomisest.

[53] Eesti Infoturbestandard. https://eits.ria.ee

7. Tulemused

Eestikeelsete kasutajate paroolide harjumuste ja mustrite uurimise tulemusena oli eesmärk luua paroolide ründesõnastik. Levinud mustreid ja sõnesid kattev andmestik, mida oleks võimalik kasutada ettevõttes või süsteemis kasutaja sisestatud parooli vastu. Osa ründesõnastiku failidest on lisatud veebirakenduse Paroolikratt andmebaasi, kus tavakasutaja saab sisestatud parooli selle vastu testida. Järgnevalt antakse ülevaade sõnastike loomisest ja paroolimurdmise tarkvara kasutamisest turvatestimise eesmärgil.

7.1 Sõnastiku koostamine

Sõnastiku koostamise peamiseks aluseks oli küsitlusuuringust saadud sisend, kus vastajad andsid oma paroolide põhjal hinnangu inspiratsiooniallikate ja mustrite osas. Kõiki valdkondi katta polnud võimalik, sest see oleks nõudnud suuremat kettamahtu, jõudlust ning sobivate lähteandmete olemasolu. Sisestatud paroolidest ja vastustest ilmnes, et põhiteemad on:

- Nimed;
- Suvalised asjad ehk lemmad;
- Hobid, mis on kaetud lemmades;
- Aadressid (pole siinjuhul kaetud, va kattuvus lemmadega).

Sõnastike genereerimise aluseks olnud algmaterjal tuli veebiallikatest:

- Eesti kirjakeele sagedussõnastik[54];
- Eesti Keele Instituudi isikunimeandmebaas[55];
- Eesti Keele Instituudi sõnaloendid[56];
- Vikipeedia mehe-[57] ja naisenimede[58] loend.

Kuna paroolides tehtud täheasendusi (nt ä -> 2) on keeruline tuvastada, siis need tehti autorile teadaolevalt levinumate asenduste ja slängi põhjal. Samas oli võimalik analüüsi tulemusena vähemlevinud sümboleid välistada, kui sümbolil oli väike esinemissagedus (nt $, £, ¤).

[54] Eesti kirjakeele sagedussõnastik. https://www.cl.ut.ee/ressursid/sagedused/index.php?lang=et
[55] Eesti Keele Instituudi isikunimeandmebaas. https://keeleabi.eki.ee/isikunimed/index.php?t=B
[56] Eesti Keele Instituudi sõnaloendid. https://www.eki.ee/tarkvara/wordlist
[57] Mehenimede loend. Vikipeedia. https://et.wikipedia.org/wiki/Mehenimede_loend
[58] Naisenimede loend. Vikipeedia. https://et.wikipedia.org/wiki/Naisenimede_loend

Mustrite genereerimisel arvestati küsitlusuuringust saadud infoga. Näiteks kasutatakse nimede puhul suur algustähte ning aastaarvu või numbri lõppu lisamist. Kuna küsitlusuuringust tuli välja ka, et peamised viimased kolm märki on ka kaks numbrit ja sümbol, siis on ka see muster kaetud. Täpsemad mustrid on lisas 1, kus on link GitHub'i repositooriumile[59].

7.2 John the Ripper'i kasutamine

Süsteemis uue paroolide turvapoliitika juurutamiseks on kaks võimalust. Esiteks saab kõiki kasutajaid suunata teatud aja jooksul oma paroole vahetama nii, et need vastaksid uutele nõudmistele. Teine võimalus on testida aktiivsete kasutajate salasõnasid, kellelt palutaks isiklikult (meili teel) või tehniliselt (keskkond suunab) oma väheturvaline parool ära vahetada.

JTR (John the Ripper[60]) on avatud lähtekoodiga tarkvara paroolirünnete teostamiseks. Alternatiivina võib kasutada ka tarkvara Hash Cat[61], kuid JTR on autorile tuttavam, sest seda kasutatakse palju CTF (*Capture the Flag*) keskkondades. Turvatestimist silmas pidades on soovitatav kasutada mõlemat tarkvara, sest need rakendavad mõnevõrra erinevaid algoritme ja seega katavad alternatiivseid mustreid [33]. Praegu kasutatav versioon on John the Ripper jumbo[62], kuid juhendi koostamisel on lähtutud ka John the Ripper repositooriumi teise haru dokumentatsioonist[63].

JTR pakub paroolide murdmiseks erinevaid meetodeid, mida on võimalik omavahel ka kombineerida [34].

- *Worldlist mode*, mille puhul saab eelgenereeritud sõnastiku ette anda.
    - `--worlist=[sõnatiku faili asukoht]`
- *Single crack mode*, mis kasutab murdmiseks teadaolevad informatsiooni (kasutajatunnust või nime). Näiteks kasutajanime *admin* proovib see *AdMiN*, *aDmIn*, *AdMin* jne.
    - `--single`
- *Incremental mode*, mis proovib ilma etteantud parameetrite või failideta luua võimalike märgikombinatsioone. Meetod on üsna kulukas ja ebaefektiivne.
    - `--incremental=[valik]`

[59] anettparismaa. GitHub. https://github.com/anettparismaa/eesti_dictionary
[60] John the Ripper password cracker. Openwall. https://www.openwall.com/john
[61] Hash Cat. GitHub. https://hashcat.net/hashcat
[62] Openwall. GitHub. https://github.com/openwall/john/tree/bleeding-jumbo/doc
[63] Openwall. GitHub. https://github.com/openwall/john/tree/e0b1de31fa92da9213c80a6fbe52c7914854cd63/doc

- *Markov mode*, sarnaneb sõnastiku meetodile, kuid on keerukam, sest sõned koostatakse algoritmi tulemusena arvestades märkide tõenäosuslikku järgnevust.
    - `--markov[=level[:algus:lõpp[:maxpikkus]]`
- *Mask mode*, kus on võimalik defineerida reeglid (mustreid), mille põhjal hakatakse salasõnu genereerima.
    - `--mask=[muster]`
- *Subset mode*, mis kombineerib etteantud märkide listi põhjal erineva pikkusega paroole, kuid on kulukas ja ebaefektiivne.
    - `--subsets=N`
- *Regex mode*, mis on võimsam, kuid kulukas ja sellele pole dokumentatsiooni. On kasulik mustrite defineerimisel, kuid seda saab asendada ka *mask mode*'iga.

Siin käsitletakse sõnastiku (*wordlist mode*) ja maskeerimise (*mask mode*) viisi koos erinevate parameetritega. Võimalikud parameetrid leiab dokumentatsioonist[64] või käsuga `john --help`. Valime otsitavaks parooliks *Kübar32!* mis räsitakse algoritmiga SHA256. Toetatavad kodeeringud leiab käsuga `john --list=formats`. Räsi kirjutatakse faili *hash.txt*, mille sisu saab olema:

```
user1:977f3ebb72235efe3bdd7b85f0ab6038841b9dcb81d02a489d092b70892ea9df
```

JTR'le on võimalik ka ainult räsidega fail (ilma ühegi parameetrita) ette anda, kuid sel juhul pole rakenduse töö eriti efektiivne. Hakatakse erinevaid räsialgoritme proovima vaikesõnastiku *password.lst* vastu ning murdmine muutub ajaliselt kulukaks ja ilmenvad jõudlusprobleemid.

Teades algoritmi saab seda ka räsifailile rakendada: `john --format=raw-sha256 hash.txt`

7.2.1 Sõnastikrünne

Testimise efektiivsemaks tegemiseks on võimalik kasutada eelgenereeritud sõnastikku, nt *yhend.txt* GitHub'i repositooriumis[65], praeguses näites nimega *generated.txt*:

```
john --format=Raw-SHA256 --wordlist=generated.txt hash.txt
```

Väljundis on näha (joonis 12), et JTR kuvab murtud parooli koos sulgudes kasutajanimega. Samuti antakse infot faili, kodeeringu ning kulunud aja kohta. Murtud paroolid salvestatakse koos räsidega *~/.john/john.pot* faili. Vahemälu saab kustutada käsuga `rm -rf /home/kali/.john`

[64] John the Ripper's command line syntax. Openwall. https://www.openwall.com/john/doc/OPTIONS.shtml
[65] anettparismaa. eesti_dictionary. GitHub. https://github.com/anettparismaa/eesti_dictionary

```
┌──(kali㉿kali)-[~/Desktop]
└─$ john --format=Raw-SHA256 --wordlist=generated.txt hash.txt
Created directory: /home/kali/.john
Using default input encoding: UTF-8
Loaded 1 password hash (Raw-SHA256 [SHA256 128/128 SSE2 4x])
Warning: poor OpenMP scalability for this hash type, consider --fork=3
Will run 3 OpenMP threads
Press 'q' or Ctrl-C to abort, almost any other key for status
Kübar32!         (user1)
1g 0:00:00:00 DONE (2024-05-11 10:45) 12.50g/s 1228Kp/s 1228Kc/s 1228KC/s krapsus..leostustega
Use the "--show --format=Raw-SHA256" options to display all of the cracked passwords reliably
Session completed.
```

Joonis 12. Käsu väljund Kali Linuxis (sõnastiku kasutamine).

Sõnastiku moodust saab kasutada ka koos reeglitega (väljund on näha joonisel 13), mis lisatakse käsurea parameetrina või */etc/john/john.conf* faili. Näiteks suurtähestamiseks kasutatakse *c*'d ja sõne lõppu lisamiseks *Az*'d:

```
[List.Rules:Example]

cAz"[0-9][0-9][0-9]"

cAz"[0-9][0-9][!.]"
```

```
┌──(kali㉿kali)-[~/Desktop]
└─$ john --format=Raw-sha256 --wordlist=generated.txt --rules=example  hash.txt
Created directory: /home/kali/.john
Using default input encoding: UTF-8
Loaded 1 password hash (Raw-SHA256 [SHA256 128/128 SSE2 4x])
Warning: poor OpenMP scalability for this hash type, consider --fork=3
Will run 3 OpenMP threads
Press 'q' or Ctrl-C to abort, almost any other key for status
Kübar32!         (user1)
1g 0:00:00:35 DONE (2024-05-11 11:12) 0.02801g/s 7358Kp/s 7358Kc/s 7358KC/s Knneski32!..Lrise32!
Use the "--show --format=Raw-SHA256" options to display all of the cracked passwords reliably
Session completed.
```

Joonis 13. Käsu väljund Kali Linuxis (reeglite kasutamine).

Rohkem reegleid on välja toodud dokumentatsioonis[66].

7.2.2 Maskeerimine

N-ö maski defineerimise eelis on see, et nii suurendab märkimisväärselt testimise efektiivsust. Järgnev näide on kohandatud Phil Bramwelli raamatust [35]. Soovitakse murda 8-märgilist parooli, mis sisaldab numbreid ning suur- ja väiketähti eesti tähestikust (koos võõrtähtedega). Ilma maskita on selleks $((2 \times 32) + 10)^8 = 899\,194\,740\,203\,776$ võimalust, kuid arvestades maski on $32 \times (2 \times 32 + 10)^5 \times 10^2 = 7\,100\,821\,196\,800$ võimalust. Seda on 99,2% vähem, mis näitab mustrite rakendamise efektiivsust.

[66] Openwall. GitHub. https://github.com/openwall/john/blob/bleeding-jumbo/doc/RULES

Eesti keele täpitähtede tõttu saab JTR'i mõningaid käske kasutada ainult kodeeringut määrates, kas käsureal või *john.conf* failis.

```
john --format=Raw-sha256 --mask=?u?b?b?b?b?d?d?s hash.txt --internal-
encoding=iso-8859-1
```

Siin defineeritud maskeeringu puhul on *?u* suurtäht; *?b* on kõik märgid (0x01-0xff); et ka täpitähed oleks hõlmatud; *?d* on number ja *?s* on sümbol.

Samuti on võimalik kasutada hübriidvariant, nt on sõnastikus lemmad ja soovides genereerida neile numbreid lõppu, saab valida:

- `john --format=Raw-sha256 --wordlist=generated.txt --incremental=digits --mask=?w?d?d?d`
- `john --format=Raw-sha256 --wordlist=generated.txt --mask='?w?d?d?d' hash.txt`

Siin asendatakse *?d* genereeritud numbritega.

7.2.3 Soolatud räsi murdmine

Kristjan Pühveli lõputöös teostatud testimine oli NTLM räside vastu [2]. Sellisel kujul räsisid leiab tänapäeval harva, sest paroolide kaitsmiseks kasutatakse enamasti soolamist. Mistõttu on ka tabelründed oma efektiivsuse kaotanud, iga räsi jaoks tuleks eraldi tabel välja arvutada. Praeguses näites arvutame paroolile (*kaabu)* räsi koos soolaga (*tugevpar00l*) kasutades algoritmi SHA256. Kasutatud on vormingut, kus paroolile lisatakse sool ja seejärel räsitakse, mille tulemusena kirjutatakse faili:

```
user1:f5cf172014df4306fad27dee802bef24fa176ae118d360251e0ae41d0521af54$tugevpar00l
```

Vorminguid saab leida käsuga `john --list=subformats | grep [algoritmi_nimi]`

Siin näites on sobiv formaat `Format = dynamic_62  type = dynamic_62: sha256($p.$s)`

Seejärel saab kasutada käsku järgnevate parameetritega:

```
john --format=dynamic_62 --wordlist=generated.txt soolatud.txt
```

Eelnenud käsu väljund on näha joonisel 14.

```
┌──(kali㉿kali)-[~/Desktop]
└─$ john --format=dynamic_62 --wordlist=generated.txt soolatud.txt
Created directory: /home/kali/.john
Using default input encoding: UTF-8
Loaded 1 password hash (dynamic_62 [sha256($p.$s) 128/128 SSE2 4x])
Warning: no OpenMP support for this hash type, consider --fork=3
Press 'q' or Ctrl-C to abort, almost any other key for status
kaabu            (user1)
1g 0:00:00:00 DONE (2024-05-11 19:09) 20.00g/s 974400p/s 974400c/s 974400C/s j nkilikega..kaadrikkudeg
Use the "--show --format=dynamic_62" options to display all of the cracked passwords reliably
Session completed.
```

Joonis 14. Käsu väljund Kali Linuxis (soolatud sisend).

Tasub märkida, et faili sellise vormistamise puhul `format=dynamic_61 type = dynamic_61: sha256($s.$p)`, kus sool eelneb paroolile, nõuab JTR räsifaili puhul sama järjestust:

```
user1:04687ac110c0e721e2004dd82848a6a082c8400518654d9d0fcf935c58657716$tugevpar001
```

Vabavaraliste tarkvarade hulgas mitmed võimalusi turvatestimiste läbiviimiseks, kuid sobiva leidmine nõuab süsteemi omapärade ja arvutusvõimuse arvestamist.

Tartu Ülikooli 2022. aastal läbi viidud turvatestimise [6] tulemusena leiti, et ca 10% kasutajate paroolid olid eestikeelsele ründesõnastikule haavatavad. Nendel kasutajatel paluti oma paroolid ära vahetada. Tulemused andsid tõuke kaheastmelise autentimise kehtestamiseks ja paroolide nõuete tugevdamiseks. Paroolide turvatestimisega on enamasti võimalik potentsiaalset kahju ennetada ning kasutajate üldist küberteadlikkust tõsta.

Kokkuvõte

Bakalaureusetöö eesmärk oli kaardistada eestikeelsete Interneti-kasutajate harjumusi ja mustreid paroolide loomisel. See hõlmas nelja suuremat tegevust, milleks olid: küsitlusuuring, lekkinud paroolide analüüs, veebirakendus paroolide testimiseks ja loomiseks ning juhend IT-spetsialistidele eestikeelsete paroolide ettevõttesiseseks testimiseks.

Küsitlusuuringus oli 225 osalejat, kellest 202 vastuseid kasutati paroolide sõnastiku koostamiseks. Tulemustest selgus, et parooligeneraatorid ja -haldurid pole vastanute igapäevaharjumustes liiga juurdunud. Samuti järgivad vastanute paroolid üldlevinud teemasid (hobid, nimed, esemed) ja mustreid (suurtäht alguses ja number lõpus). Seega saaks parooliründe ohvriks sattumise vältimiseks nende keerukust tõsta. Lisades näiteks suurtähe alguse asemel keskele, muudab see ründe läbiviimiseks loodava sõnastiku komplitseeritust. Tugevust kasvatab ka pikkuse suurendamine, kui kasutada salasõnade asemel salafraase.

Ründesõnastiku koostamise aluseks võeti lemmade ja nimede andmestik, mille põhjal genereeritud failide maht tuli kokku üle 50 GB. Nende rakendamiseks loodi veebisait Paroolikratt[67], mille andmebaasis on üle 50 miljoni kirje. Nädala jooksul laekunud 368 päringust leiti 45,9% kattuvus lekkinud või genereeritud paroolidega. Serveri andmebaasis on ka eesti domeeniga lekkinud paroolid, mis koosneb umbes 187 000 avalikust lekkest kogutud unikaalsest kirjest. Veebirakenduses on veel salafraaside generaator, millega saab luua eestikeelseid paroole.

Viimases peatükis kirjeldati sõnastiku potentsiaalseid kasutusvõimalusi, nt turvatestimine tarkvaraga John the Ripper. Peale andmeturbe kursusel antava lühiülevaate, ei õnnestunud leida juhendit eestikeelsete paroolide murdmiseks või testimiseks. Kuigi paroolidealase teadlikkuse tõus aitaks võimalikke kahjusid oluliselt vähendada.

Käesolevas bakalaureusetöös uuriti eestikeelsete paroolide mustreid ja toodi näiteid eestikeelse ründesõnastiku koostamise ning kasutamise kohta. Loodetavasti aitavad töös kajastatud näited küberturbejuhtidel oma ettevõttes turvatestimist läbi viia ning ennetada paroolidega seotud küberkuritegusid.

[67] Paroolikratt. https://paroolikratt.ee

Kasutatud allikad

[1] Küberturvalisuse aastaraamat 2024. Riigi Infosüsteemi Amet. https://www.ria.ee/sites/default/files/documents/2024-02/RIAkuberturvalisuse-aastaraamat-2024.pdf (vaadatud 21.04.2024)

[2] Pühvel, K. Paroolide murdmine ja räsifunktsioonid TÜ kasutajakontode näitel. TÜ arvutiteaduse instituudi bakalaureusetöö. 2022.

[3] Murray, H.; Malone, D. *Adaptive password guessing: learning language, nationality and dataset source. Annals of Telecommunications*, 2023, vol 78, lk 385–400, doi: 10.1007/s12243-023-00969-4.

[4] What Is Combo List? Scirge. https://scirge.com/glossary/combo-list (vaadatud 01.05.2024)

[5] Tran, L.; Nguyen, T.; Seo, C.; Kim, H.; Choi, D. A Survey on Password Guessing. 2022, doi: 10.48550/arXiv.2212.08796.

[6] Alo Peets: "Paroolide murdmine ja räsifunktsioonid TÜ kasutajakontode näitel". University of Tartu Institute of Computer Science. YouTube. 28.07.2022. https://www.youtube.com/watch?v=7MowoL1sP3c (vaadatud 14.03.2024)

[7] He, D.; Zhou, B.; Yu, H.; Cheng, Y.; Chan, S.; Zhang, M.; Guizani, N. *Group-based Password Characteristics Analysis. IEEE Network*, 2021, vol. 35, no. 1, lk 311-317, doi: 10.1109/MNET.011.2000354. https://ieeexplore.ieee.org/document/9246624 1.05.2024

[8] Taneski, V.; Heričko, M.; Brumen, B. *Analysing real students' passwords and students' passwords characteristics received from a questionnaire*. 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 2016, lk 1436-1441, doi: 10.1109/MIPRO.2016.7522365. https://ieeexplore.ieee.org/document/7522365 1.05.2024

[9] What authentication and verification methods are available in Microsoft Entra ID? Microsoft. https://learn.microsoft.com/en-us/entra/identity/authentication/concept-authentication-methods (vaadatud 09.05.2024)

[10] Brute Force Attack. Beyond Identity. https://www.beyondidentity.com/glossary/brute-force-attack (vaadatud 04.12.2023)

[11] OWASP Top Ten. OWASP. https://owasp.org/www-project-top-ten (vaadatud 31.03.2024)

[12] Hofstede, R.; Jonker, M.; Sperotto, A.; Pras, A. Flow-Based Web Application Brute-Force Attack and Compromise Detection. J Netw Syst Manage, 2017, vol 25, lk 735–758, doi: 10.1007/s10922-017-9421-4.

[13] Bošnjak, L.; Sreš, J.; B. Brumen, B. *Brute-force and dictionary attack on hashed real-world passwords. 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. Opatija, Croatia, 2018, lk 1161-1166, doi: 10.23919/MIPRO.2018.8400211.

[14] Eesti Infoturbestandard. Riigi Infosüsteemi Amet. 2023. https://eits.ria.ee (14.05.2024)

[15] Parooli loomine ja kasutamine. Riigi Infosüsteemi Amet. 02.11.2022. https://www.ria.ee/kuberturbe-nouanded/nouanded-internetikasutajale/parool (vaadatud 14.05.2024)

[16] Cost of a Data Breach Report 2023. IBM. https://www.ibm.com/downloads/cas/E3G5JMBP (vaadatud 14.05.2024)

[17] Lepik, K.; Harro-Loit, H.; Kello, K.; Linno, M.; Selg, M.; Stömpl, J. Intervjuu. Sotsiaalse Analüüsi Meetodite ja Metodoloogia õpibaas. 2014. https://samm.ut.ee/intervjuu (vaadatud 13.05.2024)

[18] Rämmer, Andu. Valimi moodustamine. Sotsiaalse Analüüsi Meetodite ja Metodoloogia õpibaas. 2014. https://samm.ut.ee/valimid (vaadatud 30.04.2024)

[19] Ettevõtte andmebaasist laeti ebaseaduslikult alla Apotheka kliendikaartide omanike andmeid. Riigi Infosüsteemi Amet. 04.04.2024. https://www.ria.ee/uudised/ettevotte-andmebaasist-laeti-ebaseaduslikult-alla-apotheka-kliendikaartide-omanike-andmeid (vaatatud 09.05.2024)

[20] Database Leaked Eek.ee 15M. 08.05.2023. https://leakbase.io/threads/database-leaked-eek-ee-15m.4597 (vaadatud 30.04.2024)

[21] What's the dark web? Google. https://support.google.com/googleone/answer/12262331?hl=en (vaadatud 09.05.2024)

[22] Hacxx Underground. *Files*. GitHub. https://github.com/hacxx-underground/Files (vaadatud 30.04.2024)

[23] Küberturvalisuse aastaraamat 2018. Riigi Infosüsteemi Amet. https://www.ria.ee/sites/default/files/documents/2022-11/RIA-kuberturvalisuse-aastaraamat-2018.pdf (vaadatud 30.04.2024)

[24] Functional and Nonfunctional Requirements: Specification and Types. Altexsoft. 30.11.2023. https://www.altexsoft.com/blog/functional-and-non-functional-requirements-specification-and-types (vaadatud 01.05.2024)

[25] What is a RESTful API? Amazon. https://aws.amazon.com/what-is/restful-api (vaadatud 05.05.2024)

[26] Kim, K. 2013. A Guide to Connecting PostgreSQL and Pythons' Fast API: From Installation to Integration. Medium. https://medium.com/@kevinkoech265/a-guide-to-connecting-postgresql-and-pythons-fast-api-from-installation-to-integration-825f875f9f7d (vaadatud 03.05.2024)

[27] Alternatives, Inspiration and Comparisons. FastAPI. https://fastapi.tiangolo.com/alternatives/#pydantic (vaadatud 03.05.2024)

[28] Features. FastAPI. https://fastapi.tiangolo.com/features (vaadatud 03.05.2024)

[29] Python FastAPI vs Flask: A Detailed Comparison. Turing. https://www.turing.com/kb/fastapi-vs-flask-a-detailed-comparison (vaadatud 03.05.2024)

[30] Peterson, R. What is PostgreSQL? Introduction, Advantages & Disadvantages. GURU99. 16.03.2024. https://www.guru99.com/introduction-postgresql.html (vaadatud 03.05.2024)

[31] Bloominari Marketing. Color Psychology in Web Design. Medium. 28.11.2018. https://medium.com/@bloominari/color-psychology-in-web-design-f60656b8f313 (vaadatud 03.05.2024)

[32] Create and use strong passwords. Microsoft. https://support.microsoft.com/en-us/windows/create-and-use-strong-passwords-c5cebb49-8c53-4f5e-2bc4-fe357ca048eb (vaadatud 10.05.2024)

[33] Liu, E.; Nakanishi, A.; Golla, M.; Cash, D.; Ur, B. Reasoning Analytically about Password-Cracking Software. *2019 IEEE Symposium on Security and Privacy (SP)*. San Francisco, CA, USA, 2019, lk 380-397, doi: 10.1109/SP.2019.00070.

[34] Openwall. *John/doc/MODES*. GitHub. 07.03.2021. https://github.com/openwall/john/blob/bleeding-jumbo/doc/MODES (vaadatud 07.05.2024)

[35] Bramwell, P. Hands-On Penetration Testing on Windows. Birmingham: Packt Publishing Ltd. 2018. https://learning.oreilly.com/library/view/hands-on-penetration-testing/9781788295666 (vaadatud 11.05.2024)

Lisa 1

Veebirakenduse lähtekood: https://github.com/anettparismaa/paroolikratt_server

Lekkinud paroolide eraldamine: https://github.com/anettparismaa/paroolid

Küsitlusuuringu ja veebilehe vastuste uurimine:
https://github.com/anettparismaa/jupyter_bakalaureus

Sõnastike loomine: https://github.com/anettparismaa/eesti_dictionary

Sõnastik: https://tartuulikool-my.sharepoint.com/:f:/g/personal/parismaa_ut_ee/ElYeHJWBGfFDqqiorKS4xPYBXBbI0jSYoOcWBD7DthKx1A?e=T2Ksfr

Küsimustik: https://tartuulikool-my.sharepoint.com/:u:/g/personal/parismaa_ut_ee/EQhqeGY9wZxGnzMvFa6q_S0BzebAKyVxZGRdS2-2CBmZPQ?e=qqFjQl

Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Anett Pärismaa**

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

   „Eestikeelsete paroolide mustrite uurimine ja ründesõnatiku koostamine“,

   mille juhendaja on **Alo Peets**,

   reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Anett Pärismaa

15.05.2024