Bacheloroppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi og elektroteknikk
Institutt for informasjonssikkerhet og kommunikasjonsteknologi

Thugitha Kanavathi
Mats D. Jensen
Martin Hyldmo
Marte Jørgensen

Anonym kommunikasjon på internett: Et dypdykk i Tor

Bacheloroppgave i Digital infrastruktur og cybersikkerhet
Veileder: Erjon Zoto
Mai 2023

NTNU
Kunnskap for en bedre verden

Thugitha Kanavathi
Mats D. Jensen
Martin Hyldmo
Marte Jørgensen

Anonym kommunikasjon på internett: Et dypdykk i Tor

Bacheloroppgave i Digital infrastruktur og cybersikkerhet
Veileder: Erjon Zoto
Mai 2023

Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi og elektroteknikk
Institutt for informasjonssikkerhet og kommunikasjonsteknologi

NTNU

Kunnskap for en bedre verden

Abstract

The dark web is a part of the Internet that requires special software to access, and Tor is one of them. "*Anonymous communication on the Internet: A Deep Dive into Tor*" aims to understand the technology behind Tor, its areas of use, and public perception. Furthermore, to develop a crawler to attempt data extraction from Tor. Preserving anonymity is imperative as all activities on the dark web are expected to be monitored. The field of study is broad and the issues the group has examined have required both a theoretical and practical approach. Given the limitations, a literature review has been conducted to provide background information and to explain the technology. In addition, a survey has been used to present the public perception. Last but not least, a crawler has been developed, where the result gave insights into the types of data existing on the dark web. The combination of these three research methods has given numerous findings. However, the most surprising discovery is the benefits of Tor, including the fight for freedom of speech and access to the free Internet.

Sammendrag

Det mørke nettet er en del av internett som det krever en spesiell programvare for å nå, og Tor er en av disse. "*Anonym kommunikasjon på internett: Et dypdykk i Tor*" handler om å forstå teknologien bak Tor, dets bruksområder og oppfatningen blant befolkningen. Videre, utvikle en søkerobot for å forsøke å høste data fra Tor. Å beholde anonymiteten er imperativt, da alt som gjøres på det mørke nettet må forventes å være overvåket. Fagområdet oppgaven ser på er stort og problemstillingene gruppen har sett på har krevd både en teoretisk og praktisk tilnærming. Gitt de begrensede rammer gruppen har hatt til rådighet har det blitt benyttet en litteraturstudie for bakgrunnsinformasjon og for å forklare hva teknologien er. I tillegg har det blitt benyttet en spørreundersøkelse som sier noe om oppfattelsen i befolkningen. Sist men ikke minst er det programmert en søkerobot, der resultatet gir et innblikk i hvilke typer data som befinner seg på det mørke nettet. Summen av disse tre forskningsmetodene har gitt mange svar. Likevel, det mest overraskende funnet er fordelene som finnes ved Tor, knyttet opp mot blant annet kampen for ytringsfrihet og tilgangen til det frie internett.

Forord

Datakriminalitet er en økende trend og ifølge ENISA sin rapport i 2022 kom det frem at de mest utbredte angrepene kan spores til trusselaktører som statsfinansierte hackergrupper, cyberkriminelle, hacktivister og hacker-for-hire aktører. Slike angrep som disse aktørene utfører vil ha svært kritiske konsekvenser for ofrene. Nevnte trusselaktører holder ofte til på det mørke nettet, og for å finne disse vil det derfor være naturlig å utforske deres habitat. Oppgaven stammer fra oppdragsgivers off. PhD, der han forsker på statstfinansierte hackergrupper , herunder deres bruk av kommunikasjon på det mørke nettet. Forskningsarbeidet utføres ved siden av et ansettelsesforhold i Digdir, regjeringens fremste verktøy for digitalisering i næringslivet.

"Anonym kommunikasjon på internett: Et dypdykk i Tor" sikter på å skape en bred forståelse og innsikt i Tor. I tillegg, å høste data som inngår i oppdragsgivers forskning, ved hjelp av den egenutviklede søkeroboten. Oppgaven har gitt oss muligheten til å bli kjent med den delen av internett som kalles det mørke nettet, og programvaren Tor som er brukt for å nå dit. Arbeidet har vært både interessant og lærerikt, samtidig som det har vært utfordrende i form av utviklingen av en søkerobot som skal klare å høste data fra Tor.

Tusen takk til Raymond Hagen og Digitaliseringsdirektoratet (Digdir) for oppdraget, veiledning samt god oppfølging og samarbeid gjennom hele arbeidet.
Tusen takk til vår veileder Erjon Zoto for veiledning gjennom arbeidet.
Avslutningsvis, tusen takk til Mathias Viken Borgersen og Mads Helland Astrup fra Nasjonalt Cyberkrimsenter (NC3) for en informativ prat som bidro til bedre innsikt i Tor og det mørke nettet.

Innhold

Figurer

Tabeller

Kapittel 1

Innledning

1.1 Bakgrunn

Dataangrep med økonomisk vinning, som løsepengeangrep og digital utpressing, har utviklet seg til å bli den typen angrep som er mest betydelig de siste årene. Dette gjelder i hele verden og det finnes mange eksempler på at trusselaktører, både statlige aktører og organiserte kriminelle, har lykkes med slike angrep. Dette vil ha svært negativ konsekvens for bedrifter eller enkeltpersoner som ender opp som ofre. Aktørene som utfører disse angrepene truer gjerne med å selge sensitiv informasjon dersom offeret ikke betaler [1]. Sensitive data som dette legges ofte ut på det mørke nettet, der mange av trusselaktørene holder til [2].

Oppdragsgiver er doktorgradsstudent på NTNU, og har i tillegg et ansettelsesforhold i Digitaliseringsdirektoratet (Digdir). Digdir er regjeringens fremste verktøy for digitalisering av næringslivet. Visjonen til Digdir er å effektivisere og samordne digitaliseringen av offentlig sektor og samfunnet [3]. Forskningen knyttet til oppdragsgivers off. PhD omhandler avanserte statsfinansierte trusselaktører, med fokus på kommunikasjonen til hackergrupper på det mørke nettet. Oppdragsgivers ønske er å avdekke mønstre og sammenhenger gjennom data lokalisert på det mørke nettet. Dette for å forsøke og gjøre en form for predikasjon, slik at det er mulig å gjøre proaktive tiltak for å hindre angrep som nevnt i forrige avsnitt.

1.2 Problemområde

Ifølge “Threat Landscape” rapporten ENISA publiserte i 2022 så kan de mest utbredte angrepene spores tilbake til trusselaktører i disse fire kategoriene: statsfinansierte aktører, cyberkriminelle, hacktivister og hacker-for-hire aktører [4]. Statsfinansierte hackergrupper har blitt mer avanserte og ressursrike i løpet av de siste årene. Det er ikke lenger bare store bedrifter og nasjoner som er målgruppen, trender viser også at små og mellomstore bedrifter, i tillegg til enkeltpersoner, er

mer utsatt enn noen gang før [4]. Dette er noe som havner i medias søkelys stadig vekk, spesielt nå i forbindelse med Ukraina-krigen og den hybride krigføringen til Russland.

Mange statsfinansierte aktører har det mørke nettet som sitt naturlige habitat [5], og illegal aktivitet har derfor satt sitt preg på nettleseren Tor helt siden lanseringen i 2002 [6]. Ifølge et studie utført av Adam K. Ghazi-Tehrani er omtrent 20 % av nettsidene som er tilgjengelig via Tor, knyttet til ulovlig aktivetet [7]. Anonymiteten som kan oppnås ved hjelp av teknologien kan være en faktor for at mange lar seg friste, men selv med et slikt verktøy er det ikke mulig å være fullstendig anonym. Med tilstrekkelig teknisk kompetanse og de rette ressursene til rådighet, er sannsynligheten for å avsløre identiteten til disse ondsinnede aktørene svært høy. Dog tilsier tallene at det ikke bare er ondsinnede aktører som benytter seg av teknologien.

I henhold til forskningsartikkelen *"The potential harms of the Tor anonymity network cluster disproportionately in free countries"* er det omtrent 6.7 % brukere globalt som benytter Tor til ulovlig formål på daglig basis. Denne prosentandelen er kun et gjennomsnitt, resultatene fra hvert enkelt land viser at tallet varierer. Basert på resultatene er det en større andel av Tor-brukere i frie land som bruker Tor til ulovlige formål, i disse landene ligger aktiviteten på omtrent 7.8 %. På den andre siden er det kun 4.8 % av all aktivitet som er ondsinnet i undertrykte land. Disse resultatene indikerer at landene som drifter infrastrukturen til Tor-nettverket, samt bistår Tor Projektet med midler, også er ofrene for de negative konsekvensene av anonymiteten Tor-nettverket sørger for [8]. Media har også hatt en viktig rolle i å forme omdømmet til denne delen av internettet. Nyhetsbildet har vært preget av assosiasjoner til illegal aktivitet og det mørke nettet. Å bli eksponert for denne tilknytningen kan føre til fordommer og redsel. Dette er hovedårsaken til at store deler av befolkning forbinder det mørke nettet med ondsinnede aktører, derav ikke valgt å utforske denne delen av internettet. For å endre dette er det viktig å sette søkelys på de positive aspektene ved teknologien.

1.3 Oppgavedefinisjon

Oppgaven sitt mål er å få en oversikt over hvilke typer data, og datamengder som ligger på Tor nettet. Oppgaven er teknisk i den forstand at gruppen må sette seg inn i hvordan Tor fungerer, og ikke minst finne, eventuelt utvikle en crawler som kan hente ned data. Det er også imperativt viktig at anonymiteten bevares når eksperimentet gjennomføres. Alt skal selvsagt dokumenteres, men alt som gjøres på det mørke nettet må forventes å være overvåket. Det sies at det finnes bare tre typer brukere på det mørke nettet. De som har noe å skjule, de som etterforsker, og sikkerhetsforskere. Gruppemedlemmenes rolle være sikkerhetsforskere. Oppgavens minstekrav:

- Forstå hvordan det mørke nettet er bygget opp
- Finne en måte å høste data fra ulike sider på det mørke nettet.
- Laste innholdet som høstes over i en database for «offline» bruk.

1.4 Avgrensning

Det er mulig å nå ulike deler av det mørke nettet ved bruk av forskjellige verktøy, denne oppgaven er avgrenset til å utforske The Onion Router (Tor), heretter kun kalt Tor. Oppgaven fokuserer videre på å opparbeide en kunnskap og oversikt over Tor, og utvikling av en søkerobot for høsting av data. Målet med dette er å gi oppdragsgiver en pekepinn på hvilke typer data som er på Tor, slik at han kan benytte informasjonen og dataene i sin egen forskning. På grunn av omfanget av oppgaven valgte gruppen å ikke fokusere på mengden data som ligger på Tor.

For å trygt kunne benytte seg av det mørke nettet vil gruppen fordype seg i hvordan Tor er bygget opp teknisk, og hvordan det fungerer. I tillegg sette seg inn i bakgrunnen og historien. Digdir har utstyrt gruppen med fire PCer som kan benyttes til forskning på Tor, da disse PCene ikke vil kunne spores til noen av gruppens medlemmer. Gruppen vil benytte seg av VPN ved bruk av Tor, for å forsikre seg om at anonymiteten er bevart under høstingseksperimentet.

Teoridelen er avgrenset til å se på fordelaktig bruk av Tor, da omfanget ville blitt for stort om gruppen skulle analysert all aktivitet som skjer på Tor.

Det er antatt at det finnes store mengder med data på Tor. For å få tak i mest mulig relevant data i forbindelse med oppdragsgivers ønsker, vil gruppen bruke tiden på å finne data som gjelder datakriminalitet.

1.5 Problemstilling

I henhold til informasjon gitt av oppdragsgiver (vedlegg C), er det dokumentert at en betydelig andel av de som engasjerer seg i ulovlige digitale aktiviteter opererer innenfor det mørke nettets grenser. Deriblant statsfiniansierte hackere som oppdragsgiver forsker på. For å kunne forstå og finne informasjon og data om disse må en utforske den delen av internett de holder til på. Tor er ett av programvarene en kan benytte for å nå det mørke nettet, og for å få en bred og detaljert forståelse av Tor må en utforske flere av dets aspekter. Det eksisterer ikke en enkelt kilde som inkluderer en helhetlig og bred forståelse av Tor på ett og samme sted, og det kan være utfordrende å navigere gjennom alle de ulike kildene som innehar varierende grad av informasjon om Tor. De fleste forbinder Tor og det mørke nettet med noe negativt og det er som regel historier med negativ vinkling som dukker opp i media. Tors historie og faktiske hensikt virker ikke å være kjent blant befolkningen, noe som kanskje er grunnen til at det hovedsaklig blir brukt til ulovlig aktivitet, da få vet om mulighetene som finnes.

1.6 Problemformulering

Målet med denne bacheloroppgaven er å få en dyp og bred forståelse av Tor, i tillegg finne en måte å høste data oppdragsgiver kan benytte i sin videre forskning. For å få til dette må det legges et kunnskapsgrunnlag om bakgrunnen og historien til Tor, samt å få en forståelse av hvordan Tor protokollen er bygd opp, og hva Tor kan benyttes til. Dette vil være det teoretiske fundamentet for oppgaven. Det vil altså bli gjennomført et nøytralt dypdykk i Tor. Videre vil gruppen også se på hvordan oppfatningen av det mørke nettet er blant befolkningen. Dette kan brukes som et ledd i å nå målet med oppgaven og knyttes opp mot fordeler som finnes med Tor. For å finne en måte å høste data trengs det først og fremst en bred forståelse av Tor i sin helhet. I tillegg, tid til navigasjon på Tor-nettleseren for å kunne utvikle et tilpasset verktøy. I denne rapporten vil gruppen dermed utføre følgende for å oppnå dette:

- Innhente informasjon som legger til grunn for det teoretiske fundamentet for oppgaven.
- Gjennomføre et lite systematisk litteraturstudie
- Gjennomføre en spørreundersøkelse for å se på hvordan det mørke nettet blir oppfattet
- Utvikle en søkerobot for høsting av data

1.7 Prosjektmål

1.7.1 Effektmål

- Få oversikt over oppbyggingen, bakgrunnen og historien til Tor, samt data som ligger der.
- Utvikle en søkerobot som høster dataene nevnt over.
- Få hentet ut mest mulig informasjon om hackergrupper fra det mørke nettet via Tor, slik at det kan forskes videre på.

1.7.2 Resultatmål

- Ha en ferdigutviklet søkerobot som kan hente ned innhold fra det mørke nettet relatert til nøkkelord en vil lete etter.
- Å ha en leveringsklar og oversiktlig database med ulike data fra det mørke nettet.
- Gjennomføre bacheloroppgaven på en sikker måte, slik at ingen på gruppen blir eksponert på det mørke nettet.

1.7.3 Læringsmål

- Opparbeide kunnskap om det mørke nettet, Tor, hvordan det fungerer og henger sammen.

- Opparbeide kunnskap om hvordan en kan holde seg sikker og anonym på det mørke nettet.
- Gjennomføre et større prosjektarbeid og metodisk arbeid.
- Hvordan lage en søkerobot og hvordan den fungerer.

1.8 Målgruppe

Målgruppen for datahøstingen i oppgaven vil være oppdragsgiver i den grad at gruppens resultater vil bli brukt i videre forskning. Videre er målgruppen for rapporten både norske og andre skandinaviske akademikere, samt medstudenter med interesse for fagfeltet. Rapporten skal kunne gi god innføring i hva teknologien er, og derfor vil også hele eller deler av rapporten kunne benyttes av Digdir til opplæring, noe som kan resultere i økt bevissthet rundt temaet. Oppdragsgiver vil, i tillegg til å være i målgruppen for selve datahøstingen, befinne seg innenfor målgruppen for denne rapporten, ettersom den også kan bli brukt i hans videre forskning. Det blir brukt en del IT-terminologi i rapporten, dersom leseren ikke har forkunnskaper innenfor fagfeltet er det anbefalt å ta en gjennomgang av ordlista i forkant.

1.9 Prosjektgruppens bakgrunn

Prosjektgruppen går på siste året av en bachelorgrad i *Digital Infrastruktur og Cybersikkerhet*, ved Norges teknisk-naturvitenskapelige universitet på campus Gjøvik. Gjennom studieløpet til gruppen er det tilegnet kunnskap og erfaring innen områdene: cybersikkerhet og teamarbeid, datakommunikasjon og nettverk, infrastruktur, datamodellering og databasesystemer, og programmering. Disse områdene er høyst relevant for hvordan oppgaven kan løses på en god måte. Både for hvordan søkeroboten skal lages og hvordan informasjon om det mørke nettet kan formidles på best mulig måte. Sikkerhet er mye mer enn bare teknologi, det er også psykologi og sosialantropologi.

1.9.1 Hva måtte læres?

Gruppen har vært igjennom mange emner i løpet av disse tre årene, som nevnt i forrige avsnitt. Likevel har ikke det mørke nettet eller Tor vært noe tema i disse emnene og det var heller ikke noen av gruppemedlemmene som hadde erfaring med Tor eller det mørket nettet generelt fra før. På grunn av begrenset kunnskap til tema, var det mye som måtte læres før gruppen kom ordentlig i gang med oppgaven. Det gjaldt å sette seg inn i det tekniske ved Tor, slik at gruppemedlemmene hadde nok kunnskap om for eksempel hvordan holde seg helt anonym i prosessen med å høste data. I tillegg ble det brukt mange timer på navigering på Tor for å finne lenker som kunne brukes ved kjøring av søkeroboten. Å søke etter noe på Tor fungerer ikke på samme måte som ved et Googlesøk, det er mye mer tungvint,

noe som gjorde det ekstra ressurskrevende. Videre, forståelse for hvordan pakker rutes gjennom nettverket og dets innebygde sikkerhetsmekanismer, er avgjørende for å kunne programmere en søkerobot som klarer å navigere rundt på egenhånd.

1.10 Rammer

1.10.1 Tidsmessige rammer

For at gruppen skal kunne opprettholde de sikkerhetsmessige rammene forklart i avsnitt 1.10.3, er det nødvendig å opparbeide en oversikt over Tors oppbygging før utviklingen av søkeroboten kan starte. Dermed vil data som hentes fra Tor kun være det som er hentet fra søkeroboten er ferdigutviklet, til og med 10. mai.
Bacheloroppgaven skal leveres innen 22.mai 2023.

1.10.2 Tekniske rammer

Gruppen har fått tildelt fire Lenovo-PCer som var tilbakestilt til fabrikkinnstillinger før de ble gitt ut. Oppdragsgiver leverte også ut en harddisk slik at data som ble høstet kunne lagres eksternt. Dataene som lagres på denne harddisken må kunne aksesseres fra både Windows og Linux.
Rapporten blir skrevet i Overleaf i Latex. Et digitalt verktøy der gruppen kan samarbeide med skrivingen samtidig.

1.10.3 Sikkerhetsmessige rammer

Gjennom oppgaven skal gruppemedlemmene praktisere gode sikkerhetsrelaterte vaner som opprettholder anonymitet og konfidensialitet. Dette gjøres ved bruk av en VPN-tjeneste, beskrevet i avsnitt 2.6.3.

1.11 Rapportstruktur

Kapittel 2 inneholder nødvendig bakgrunnsteori for å kunne forstå oppgavens helhet, samt teori som underbygger problemstillingen presentert i avsnitt 1.5. Videre beskrives metodikken som er brukt for å løse oppgaven i kapittel 3. Resultatene fra spørreundersøkelsen og datahøstingen presenteres i kapittel 4. I kapittel 5 vil resultatene diskuteres og sammenlignes, i tillegg vil blant annet erfaringer og refleksjoner fremlegges. Kapittel 6 inneholder avslutningen, der problemstillingen og problemformuleringen vil konkluderes. Til slutt er det vedlagt relevante vedlegg, som ordliste, prosjektavtale, prosjektplan, møtereferat, gantt diagram og kildekode til søkeroboten.

1.11.1 Språk på rapporten

Rapporten er skrevet på norsk. Da dette er tilfellet, har gruppen valgt å holde seg til norske ord og uttrykk over bruk av engelske faguttrykk. I tilfeller der engelske faguttrykk ikke har en norsk oversettelse, eller ikke er mulig å fornorske, har gruppen valgt å bruke det engelske uttrykket. Disse ligger også i ordlisten. I denne oppgaven er Tor skrevet med stor forbokstav og resten små, som et egennavn. Dette er også måten Tor Prosjektet skriver det, og gruppen valgte å forholde seg til det.

1.11.2 Bruk av fet og kursiv skrift

Direkte sitat er strukturert ved bruk av kursiv, innrykk, anførselstegn og fotnote til kilden hvor sitatet er hentet. Videre er kursiv brukt gjennom oppgaven der det er hensiktsmessig for å øke lesbarheten.

Kapittel 2

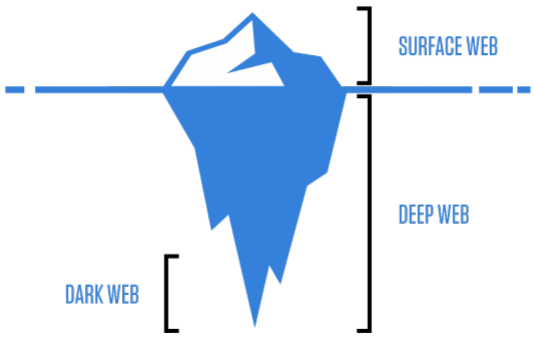
Teori

2.1 Introduksjon

Hovedmålet med dette kapitlet er å få frem det teoretiske fundamentet for oppgaven, som er med å bidra til å skape en bred forståelse av Tor, som beskrevet i problemformuleringen avsnitt 1.6. Innholdet er oversiktlig delt inn i fem deler. Først presenteres Tors bakgrunn og historie, før det så vil legges vekt på positive aspekter ved bruken av Tor, samt relatert arbeid. Deretter er det innhentet informasjon om teknisk oppbygging med fokus på de verktøyene som gjør det mulig å bevare anonymiteten på det mørke nettet. Til slutt blir det gjort rede for nødvendig bakgrunnsinformasjon i forbindelse med utviklingen av søkeroboten. Alle disse delene vil tilsammen legge det teoretiske fundamentet for oppgaven.

2.1.1 Det åpne, det dype og det mørke nettet

Internett er et stort globalt nettverk som er satt sammen av mange mindre nettverk [9], som igjen består av mange millioner databaser, servere og nettsider. De aller fleste i dagens samfunn er og har vært på internett. Den norske befolkningen kan brukes som et eksempel. Grunnet digitalisering er det svært nødvendig med grunnleggende ferdigheter for å kunne utføre diverse gjøremål i hverdagen, og det folk flest forbinder med internett er det som kalles det åpne nettet. For å skape et visuelt bilde kan isfjell-analogien brukes, se figur 2.1.



Figur 2.1: Oversikt over inndelingen av internett i form av et isfjell[1]

Det åpne nettet

Det åpne nettet vil si nettsteder som er å finne ved hjelp av tradisjonelle nettelesere og søkemotorer som Google, Yahoo og Bing. Det åpne nettet kan derfor anses som toppen av isfjellet som ligger på overflaten, og disse nettstedene utgjør kun omtrent 5 % av hele internett [10]. Eksempler på nettsteder som dette er Wikipedia, Store Norske leksikon, eller nettaviser uten betalingsmur.

Det dype nettet

Den delen av isfjellet som befinner seg under vann kalles det dype nettet, og består av over 90 % av alle nettsteder. Størrelsen på det dype nettet gjør det vanskelig å kartlegge hvor mange nettsteder som er aktive til enhver tid. Det har blitt avklart at store deler av det dype nettet innbefatter databaser med både offentlige og private filsamlinger, som kun er søkbare inne i selve databasen. Dette inkluderer også interne nettverk (Intranett), som brukes til kommunikasjon og kontroll av aspekter innad i organisasjoner. Slike Intranett er ofte laget for en utdanningsinstitusjon, foretak eller myndigheter [10].

Andre nettsteder som befinner seg på det dype nettet kan ha ulike former for sikkerhetssperrer, passordbeskyttelse er en av dem. Innlogging i nettbanken, e-posttjenester og andre sosiale medier er eksempler på nettsteder som er passordbeskyttet. Disse nettstedene er ikke tilgjengelig direkte gjennom søkemotorer, det vil si at en når bankens nettside via søkemotorene, men ikke din personlige informasjon der. Nettsteder som befinner seg på det dype nett har et ønske om å ivareta personvern og beskytte brukerinformasjon, og derfor skjules innholdet fra det åpne nett[10].

[1] `https://www.networkboxusa.com/what-is-the-dark-web/`

Det mørke nettet

Det dype nettet omfatter også den delen som kalles det mørke nettet, men begrepsfeil i media har skapt forvirring og mange tror at dette er det samme, noe som ikke stemmer helt. Det mørke nettet er kun tilgjengelig via et lite utvalg av nettlesere og gir tilgang til nettsteder som ikke er indekserte. Denne delen av internett kan sammenlignes med tuppen av isfjellet i figur 2.1, som ligger nedest under vann, og ikke er synlig med mindre det oppsøkes aktivt. Her blir blant annet trafikken anonymisert og det er ikke mulig å utføre søk på lik linje som på det åpne nettet[10].

Den første formen for det mørke nettet ble introdusert til verden i mars 2000 av Ian Clarke, en irsk student. Programvaren ble kalt Freenet og var starten på noe som skulle bli elsket og fryktet av mange. Clarke utviklet og lanserte en programvare som tidligere hadde vært en del av et skoleprojekt. Programmet fikk navnet Freenet og var en platform som kunne tilby brukerne muligheten til å kommunisere anonymt online gjennom et desentralisert distribuert datalager som oppbevarte og leverte informasjon. Likevel fikk ikke det mørke nettet sitt gjennombrudd før "The Onion Router" (Tor) ble lansert [11].

2.2 Tor: bakgrunn og historie

Tor står for "The Onion Router" og er et gratis spesialprogram som muliggjør anonym kommunikasjon og trafikk på nettet. Det norske ordet for denne metoden er løkruting og ble utviklet på 90-tallet, men Tor i seg selv ble lansert i 2002. Prinsippet løkruting ble utviklet av matematikeren Paul Syverson og informatikerne Michael G. Reed og David Goldschlag som jobbet i forskningsavdelingen under det amerikanske marineforsvaret (The United States Naval Research Laboratory (NRL)). Bakgrunnen for at dette ble utviklet var for å muliggjøre hemmelig kommunikasjon for Amerikansk etterretning på nettet [12].

Det var informatikerne Roger Dingledine og Nick Mathewson sammen med Syverson fra NRL som utviklet en alfaversjon av "The Onion Router" med akronymet Tor. Tor ble lansert internt i 2002, men ble publisert for det offentlige allerede i 2003. I løpet av 2003 hadde nettverket i overkant av 10 noder som var satt opp av frivillige. Disse befant seg i hovedsak i USA, men det var også én i Tyskland. I starten, etter den offentlige lanseringen, ble Tor mest brukt av aktivister og teknologifolk som visste hva de drev med. For de som ikke var så teknologisk anlagt, var ikke bruken av Tor like enkel. I 2004 ble selve koden lagt ut på det åpne nettet. Utviklerne Dingledine og Mathewson fikk støtte fra Electronic Frontier Foundation (EFF) for å videreutvikle Tor, og i 2006 ble "The Tor Project" opprettet, heretter kalt Tor Prosjektet (som ikke må forveksles med Tor som er selve nettverket). De jobber for å opprettholde driften til Tor og utviklingen av det, og er i overkant av

100 ansatte over hele verden. Fra 2008 ble det utviklet verktøy som gjorde Tor mer tilgjengelig for allmennheten og aktivistene som hadde ønske om å ta det i bruk [6].

2.3 Utforskning av fordeler med Tor

Dette kapitlet utforsker årsaker til at Tor er en fornuftig løsning. Selv om Tor ofte er assosiert med den ulovlige aktiviteten som er utbredt på nettverket, er det viktig å anerkjenne at Tor også brukes av mange personer med legitime formål [13]. Dette inkluderer de som ønsker å holde sin internettrafikk privat fra nettsider og reklamebyråer, de som er bekymret for spionasje, aktivister, journalister og militært personell [12]. Det er derfor feil å betrakte Tor som en utelukkende kriminell plattform, og det er nødvendig å skille mellom ulovlig og legitim bruk av nettverket.

Oppdraget til Tor Prosjektet, som de selv har uttalt, er:

> *"To advance human rights and freedoms by creating and deploying free and open source anonymity and privacy technologies, supporting their unrestricted availability and use, and furthering their scientific and popular understanding"* [2]

2.3.1 Eksempler på hvordan TOR protokollen har bistått i situasjoner der internettsensur er utstrakt

Boken "*Hands-On Dark Web Analysis: Learn What Goes on in the Dark Web, and How to Work with It*" som kom ut i 2018 er skrevet av Sion Retzkin. Han er en IT-ekspert med over 20 års erfaring. Han bedriver opplæring innen IT-sikkerhet, etisk hacking og informasjonssystemer. Kapittel 9 i denne boka "*What Goes on in the Dark Web - Case Studies*" gir flere eksempler på hvordan det mørke nettet blir brukt, hvem brukerne er og hvorfor de er der. Det kommer tydelig frem at det finnes mange kriminelle på det mørke nettet. Det er salg av narkotika, våpen, menneskehandel, pedofile nettverk med mer, men det må også frem at det finnes mange som bruker Tor til legitime formål.

Mange land og myndigheter sensurerer deler av internettet innenfor landegrensene sine, som gjør at innbyggerne blant annet ikke får tilgang til korrekt informasjon, og heller ikke kan ytre sine meninger fritt på nett eller i sosiale media. I noen land som blant annet Kina eller Russland vil det å ytre seg negativt om regimet kunne føre til fengsel eller i verste fall få fatale konsekvenser. Det at nyhetsaviser, diverse forum og Facebook befinner seg på Tor, gjør at befolkningen kan få tilgang til informasjon uten å bli forhindret av sensur. Videre, er for eksempel menneskerettighetsaktivister i mange land anbefalt å bruke det mørke nettet

[2] `https://www.torproject.org/about/history/`

for å beskytte seg selv når de kommuniserer ut informasjon eller med hverandre. Dette er eksempler som underbygger de positive sidene ved Tor.

Global Voices er et nettsted på Tor som publiserer nyheter som ikke kommer ut i så mange kommersielle nyhetskanaler. Disse nyhetene blir oversatt til veldig mange språk for å nå ut til flest mulig mennesker, slik at de ikke blir hindret av språkbarrieren. Her er det bloggere, journalister, oversettere, aktivister og mange fler som bidrar [14].

SecureDrop

Flere nyhetssider har en tjeneste på Tor som heter SecureDrop. Dette er et verktøy som sørger for at mennesker kan sende meldinger og dokumenter til nyhetssidene helt anonymt. All kommunikasjon er kryptert. For å unngå overvåking og avlytting, blir gjerne det som leveres dekryptert på en enhet som er isolert fra internett og interne nettverk hos nyhetsavisen. Nettavisene som tilbyr denne tjenesten har sin egen .onion side på Tor [15]. Dette kalles en whistleblower-tjeneste. Det vil si at dersom det er ønskelig å rapportere om feil eller ulovligheter som skjer, for eksempel på jobb, eller andre situasjoner der en vil tipse aviser om hendelser de bør undersøke.

Den arabiske våren

Et annet konkret eksempel er den arabiske våren. Den arabiske våren er fellesbetegnelsen som ble brukt på opprørene som skjedde i flere land både i Nord-Afrika og i Midtøsten i perioden 2010-2011. Disse hendelsene førte blant annet til regimeendring i Tunisia, Egypt og Libya. Bakgrunnen for opprørene var befolkningens krav om demokratiske rettigheter og de økonomiske forholdene, og bølgen spredte seg fort [16]. Den arabiske våren er et godt eksempel på hvorfor tjenester som Tor kan være viktig. Flere av regimene fjernet innbyggernes mulighet til å koble seg på internett, og fratok de dermed muligheten til å avtale demonstrasjoner og deling av nyheter. Tor gjorde det mulig for befolkningen å få tilgang til kritiske ressurser, sosiale media og nettsteder som myndighetene hadde blokkert [6].

For å ta Egypt som eksempel valgte myndighetene den 25. januar i 2011 å blokkere Twitter, dagen etter ble også Facebook tatt ned. Allerede den 27.januar valgte de å ta ned hele internettet i Egypt, og muligheten for å sende SMS ble blokkert. WhatsApp, som ble brukt av mange som et alternativ til SMS, ble også forsøkt tatt ned. Dette gjorde at Egyptiske myndigheter hadde full kontroll på informasjon som ble sendt ut både innenlands og utenlands. Alle internettleverandører i Egypt på denne tiden var eid eller styrt av myndighetene, og leverandørene hadde dermed ikke annet valg enn å skru av. Dette gjorde at egypterne måtte finne andre metoder å kommunisere på og mange trekte ut i gatene for å oppsøke og finne informasjon om hva som skjedde [17].

Lignende hendelser skjedde også i flere andre land, som Libya, Syria, Tunisia og Saudi Arabia. I enkelte land ble det utført en rekke hackingangrep mot

motstandere av regimet, noen fikk Facebookbrukeren sin slettet og andre land overvåket brukere og aktiviteten deres. I tillegg ble det utført Denial of Service (DoS) angrep mot nettstedene til motstanderne av myndighetene [18].

Videre ble Tor og det mørke nettet også brukt til anonym kommunikasjon mellom mennesker, hvor innholdet i meldingene kunne ha ført til arrestasjon. Flere av disse landene verken har eller har hatt en ytringsfrihet på lik linje som for eksempel i Norge. Den norske befolkningen kan kritisere både regjeringen, politiet eller forsvaret uten å være redd for rettsforfølgelse eller lignende konsekvenser, noe som ikke var tilfellet under den arabiske våren [19].

Sensureringen og blokkingen av internett i Russland

Enda et eksempel på Tor Prosjektets selvpålagte oppdrag i kampen mot internettsensur, er deres løsning mot russiske myndigheters sensurering. Roskomnadzor er Russlands organ for overvåkning av informasjonsteknologi, kommunikasjon og massemedia [20]. I 2021 begynte Roskomnadzor å sensurere innholdet innbyggerne hadde tilgang til på internett ved å stramme inn VPN-tilgangen til russiske borgere [21]. Russiske myndigheter ga ordre om at internettleverandører skulle blokkere diverse tjenester, og økte gradvis sitt grep om fri tilgang til informasjon. Desember 2021 nådde bølgen av sensur Tor. Tor Prosjektets nettside ble blokkert og Tor-nettverket var ikke lenger tilgjengelig for russiske borgere. Da Tor Prosjektet fikk nyss i situasjonen ble det satt i gang et prosjekt for å finne ut hvordan russiske nettverksleverandører blokkerer Tor. Deretter ble det laget et system for å unngå disse blokkeringene. Dette systemet ble kalt "snowflake" [22], og er innebygd i Tor-nettleseren. Denne funksjonen ble utviklet i all hastverk for å kunne hjelpe russerne med å være tilkoblet til det frie internettet.

Russland er et av de landene med flest Tor-brukere, og i 2021 utgjorde russerne 15 % av totale daglige brukere [21]. Denne kommentaren sier noe om den sentrale rollen Tor har, når det gjelder russernes tilknytning til resten av verden:

> *"Tor helped me a lot. Here in Russia, blocking on the Internet is extremely common... Tor helps me bypass blocking and get more privacy. For example, many wonderful websites, such as foreign services or the websites of the Russian opposition, have been blocked. I have been using Tor for many years... without it, many very important sources of useful information would be inaccessible, or accessible with great difficulty."* [3]

2.3.2 Hvordan TOR prosjektet blir videreutviklet og vedlikeholdt

Noe som gjør Tor såpass fritt er at det blir opprettholdt av frivillighet i stor grad. Tor Prosjektet har mange ansatte som jobber med drift og utvikling, men prosjek-

[3] https://www.torproject.org/static/findoc/2020-2021-TorProject-Annual-Report.pdf

tet er avhengig av frivillige. Tor er ikke utviklet for å tjene penger, noe som gjør at myndigheter og rettsvesenet ikke har styring eller påvirkning på tjenesten Tor [23].

Doktorgradstudent Hsiao-Ying Huang og professor i informasjonssikkerhet Masooda Bashir ved Universitetet i Illinois, har gjennomført en studie utgitt i 2016 "*The onion router: Understanding a privacy enhancing technology community*". Studien er en del av "*Proceedings of the ASIST Annual Meeting, 2016, Vol.53 (1), p.1-10*" [24]. De har gjennomført en undersøkelse blant de frivillige som drifter Tor-noder rundt omkring i verden. For å blant annet finne deres motivasjon, og deres perspektiver på tema som omhandler Tor og personvern. Videre tar Huang og Bashir opp at dagens teknologi har gjort det mulig å kommunisere og knytte nettverk med andre mennesker, og dele sitt daglige liv med resten av verden. Det er mye positivt med denne utviklingen, på den andre siden finnes det også negative sider ved dette. Det samles, prosesseres og deles mye informasjon om brukerne. For eksempel ved netthandel og ved søking på Google eller andre søkemotorer. Dette fører til personlig reklame på blant annet sosiale medier eller digitale nyhetsaviser.

Det har blitt utviklet teknologier som skal beskytte personvernet og privatlivet på nett, kalt Privacy Enhancing Technologies (PET). Den mest kjente er Tor, og det er ca. 7000 noder spredt rundt i verden. Antall frivillige er ikke tallfestet, men det er mange færre enn antall noder som finnes. Det var 50 deltakere som svarte på Huangs og Bashirs spørreundersøkelse. Resultatet viste at motivasjonen til frivillige var personlige belønninger, samt ønsket om å forbedre Tor-nettverket, gleden av å løse problemet med personvern og privatliv på nett, og at de støtter verdiene som ligger bak et godt personvern. Tor gjør mye for å sørge for brukernes personvern, men stigmaet av Tor-nettverket hindrer utviklingen og tilpasningen av PETs i hverdagen [24].

2.3.3 NSA-overvåkning og Tors positive rolle i informasjonsdeling

I 2013 ble det avdekket at etterretningsorganisasjonen National Security Agency (NSA) i USA hadde overvåket både nasjonal og internasjonal tele- og datatrafikk uten å innhente tillatelse [25]. Informasjon knyttet til Amerikansk spionasje rettet mot EU og Kina ble også offentliggjort som en del av dette. Denne informasjonen så dagens lys takket være den amerikanske varsleren Edward Snowden. Snowden har en fortid som datatekniker i Central Intelligence Agency (CIA) og som digital infrastrukturanalytiker for IT-selskapet Booz Allen Hamilton. I forbindelse med dette ansettelsesforholdet fikk han også muligheten til å utføre oppdrag for NSA [26], herunder bidra til å drifte overvåkningsprogrammet kalt "XKeyscore". Dette programmet var i stand til å innhente informasjon om internettbrukere over hele verden. Informasjon som e-postdresse, samtaler og søkehistorikk ble deret-

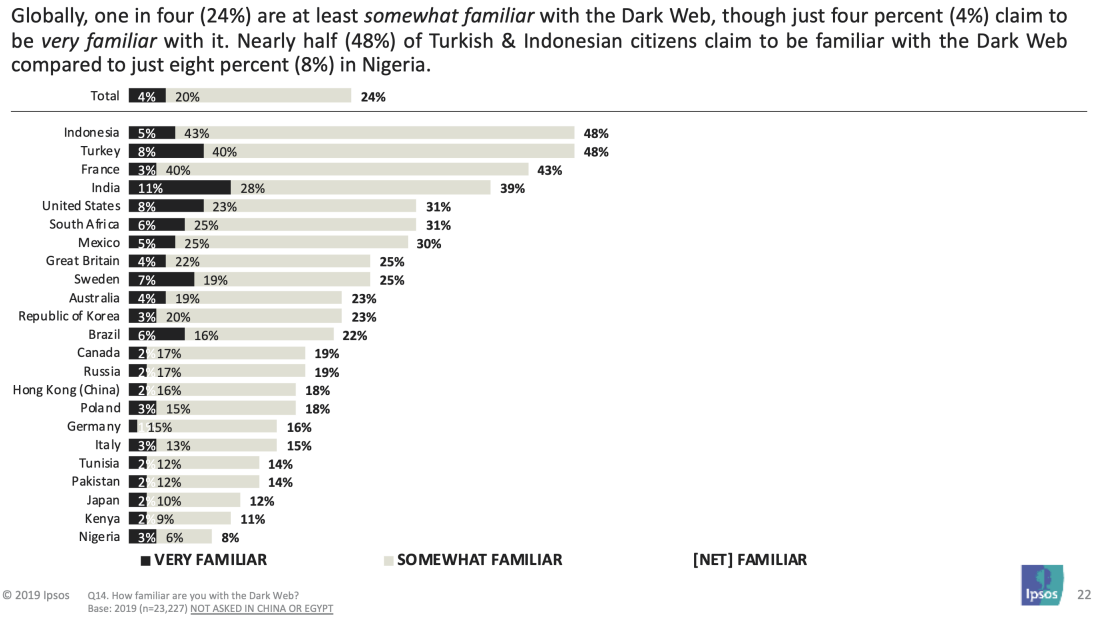
ter lagret i egne databaser, som var søkbare [27]. Søket kunne gjennomføres ved å fylle inn et skjema og dette krevde ikke noe form for godkjennelse verken fra overordnede i NSA eller en domstol [28].

Snowden brukte det mørke nettet til å lekke informasjon om NSA og XKeyscore for å forsikre seg om at nyheten nådde ut til alle verdensdeler. Han tok også kontakt med ulike nyhetsaviser ved hjelp av Tor, blant annet The Guardian for å varsle om overvåkningsprogrammet og opplyse verdensbefolkningen om bruddet på personvern. Det mørke nettet har vist seg å være et viktig verktøy for slike varslere, og på bakgrunn av dette har CIA selv lansert en Tor-basert side for å motta anonyme tips og avsløringer [29]. Dette er enda et eksempel på positive aspekter ved Tor, og hvordan et slikt verktøy kan bidra til informasjonsdeling.

2.4 Relatert arbeid

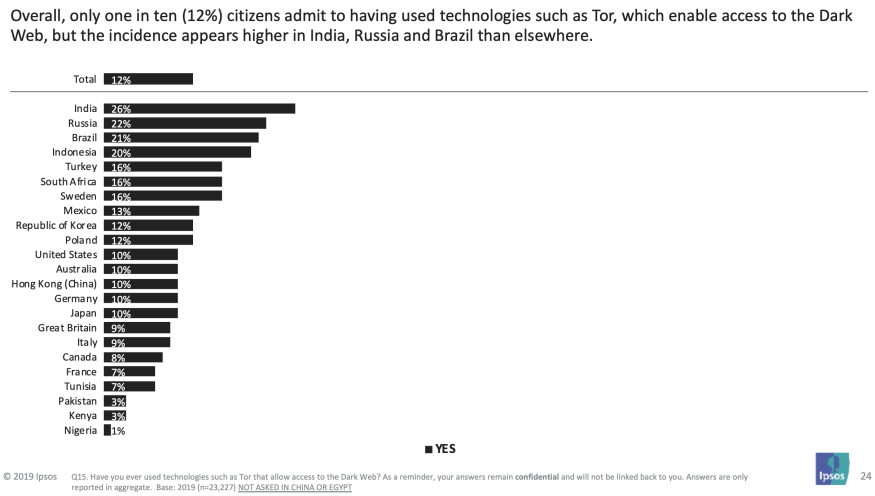
2.4.1 Spørreundersøkelse utført av CIGI

I 2019 utførte CIGI en spørreundersøkelse om det mørke nettet på verdensbasis, og mottok 23.227 besvarelser. *Centre for International Governance Innovation* er en uavhengig og partipolitisk organisasjon som har stort fokus på forskning og fremme politisk debatt [30]. Resultatene fra spørreundersøkelsen til CIGI tilsier at det kun er 24 % av verdensbefolkningen, altså 1 av 4 i verden, som har noe kjennskap til det mørke nettet [31]. Av disse 24 % er det kun 4 % av deltakerne som er godt kjent med denne delen av internett [31]. Figur 2.2 presenterer resultatene fra landene individuelt samt gjennomsnittet øverst.



Figur 2.2: How familiar are you with the dark web?[4]

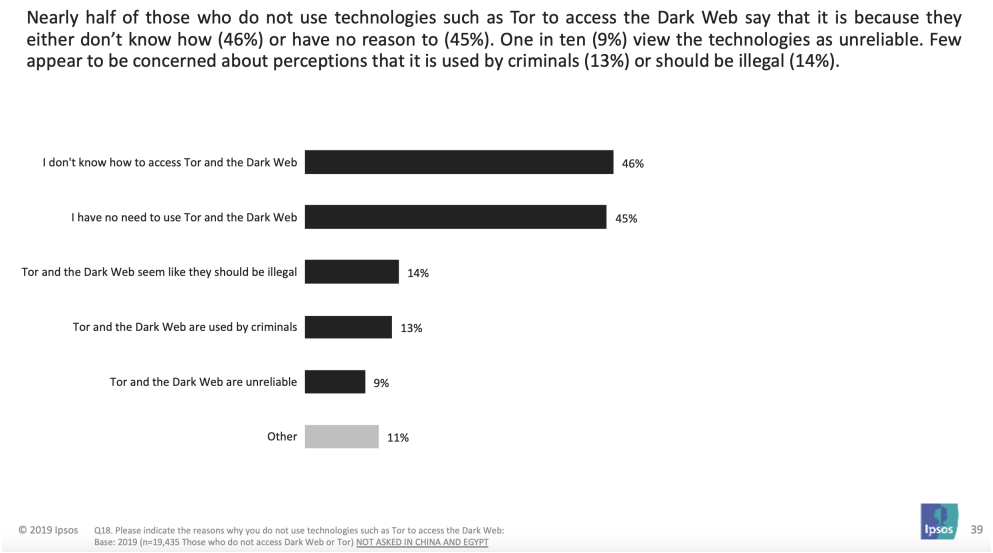
Deretter ble det spurt om deltakerne hadde benyttet seg av verktøy, som blant annet Tor, for å få tilgang til det mørke nettet. Resultatet, som vist i figur 2.3 tilsier at det kun er 12 % av deltakerne som har utforsket det mørke nettet [31].



Figur 2.3: have you ever used technologies such as Tor that allow access to the Dark web?[5]

[4] https://www.cigionline.org/sites/default/files/documents/2019%20CIGI-Ipsos%20Global%20Survey%20-%20Part%205%20Cryptocurrencies%2C%20Blockchain%2C%20Dark%20Web%20%26%20Product%20Certification.pdf

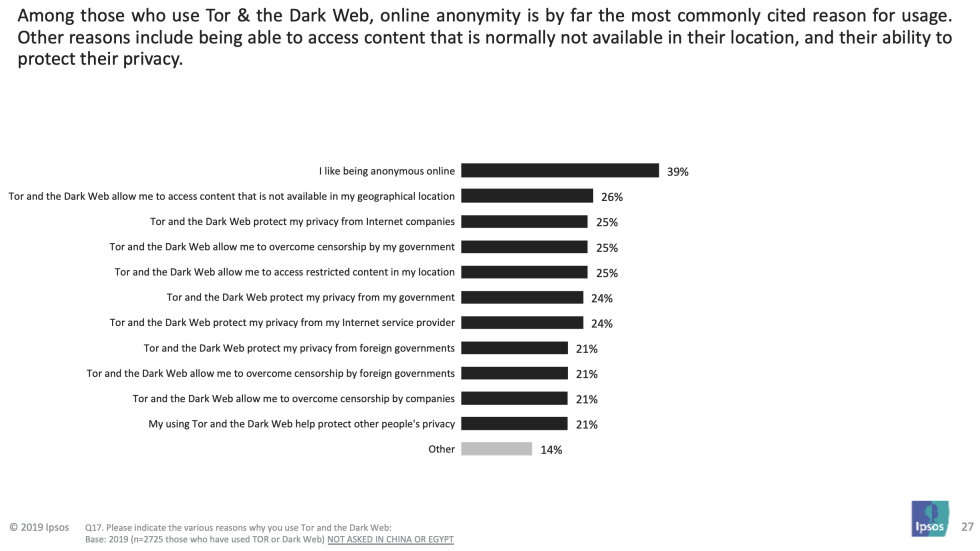
I den følgende studien ble deltakerne bedt om å svare på spørsmål angående manglende bruk av Tor og det mørke nettet. Ved hjelp av resultatene, som vises i figur 2.4, er det mulig å se at 46 % av deltakerne ikke brukte verktøyet grunnet begrenset kunnskap og ferdigheter, samt usikkerhet om hvordan en får tilgang til denne delen av internett. Videre hevdet 45 % av deltakerne at funksjonaliteten til Tor og det mørke nettet ikke var nødvendig i deres daglige liv. I tillegg, mente 27 % av deltakerne at det mørke nettet enten burde være ulovlig eller kun blir benyttet av kriminelle [31].



Figur 2.4: Please indicate the reasons why you do not use technologies such as Tor to access the dark web[6]

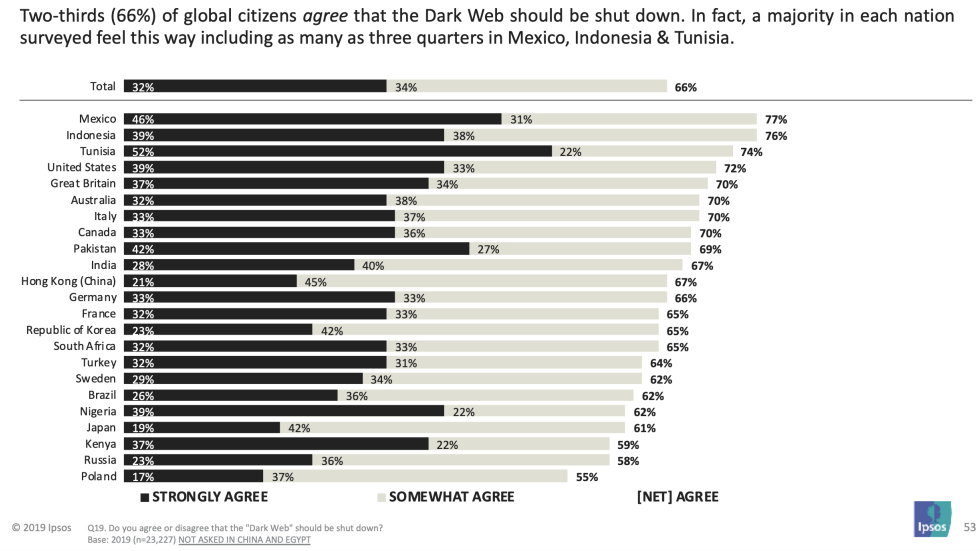
I følge CIGI, som også fremlegges i figur 2.5, er det 39 % som bruker Tor, eller som er på det mørke nettet for anonymiteten som er mulig å oppnå på denne siden av internett [31]. Videre forteller resultatene til CIGI at 26 % av deltakerne bruker Tor eller det mørke nettet til å omgå geografiske begrensninger, og få tilgang til innhold som ellers ville vært utilgjengelig [31].

[5]Se Fotnote 4
[6]Se Fotnote 4



Figur 2.5: Please indicate the various reasons why you use Tor and the Dark Web.[7]

Figur 2.6 viser at 66 % av verdensbefolkningen mener at det mørke nettet bør legges ned [31].



Figur 2.6: Do you agree or disagree that the dark web should be shut down?[8]

[7] Se Fotnote 4
[8] Se Fotnote 4

2.5 Teknisk oppbygging av Tor

Tor er ett av verktøyene som har blitt bygd på løk-ruting prinsippet, dette prinsippet er beskrevet i avsnitt 2.5.3. Nettleseren til Tor gjør det mulig for brukere å surfe på internett, men likevel forbli anonyme. Nettverket består av mange frivillige noder, over hele verden, som deler sin båndbredde og ressurser for å hjelpe brukere med å bevare anonymiteten sin, og sørge for fri tilgang til internett uten sensurering. Kombinasjonen av løkruting og utgangsnoder gjør det utfordrende å spore opp brukerens lokasjon, samt hvilken informasjon som utveksles, men det er ikke umulig. Infrastrukturen til nettverket består av tre type noder; inngangsnode, mellomnode og utgangsnode.

2.5.1 Noder

Inngangsnoder

Den første noden som brukeren kobler seg opp mot for å surfe anonymt på internett kalles en inngangsnode. Disse nodene blir valgt ved første oppstart av Tor-nettleseren. En av grunnene til at flere velger å bruke Tor er muligheten til å holde seg anonyme. Dersom en angriper er eier av både inngangsnoden og utgangsnoden som brukes, vil angriperen ha mulighet til å profilere brukeren ut fra trafikken.

Tor Prosjektet har konkludert med at for de fleste brukere vil en profilering være like ille som å bli overvåket hele tiden [32]. Ut fra denne konklusjonen er det bestemt at Tor-nettleseren vil forholde seg til tre faste inngangsnoder i fire til åtte uker før de byttes ut. Dersom minst to av disse er nede på samme tid, vil Tor-nettleseren finne nye midlertidige inngangsnoder som vil brukes [33]. Det er iverksatt tiltak for å gjøre angrep via inngangsnoder mer tidkrevende. Et av de er at noder som nylig er satt opp ikke vil kunne bli brukt som inngangsnoder. For at en node skal kunne bli valgt ut som en inngangsnode må den oppfylle krav innen båndbredde, oppetid og levetid. Når Tor-nettleseren har valgt ut de tre inngangsnodene sine, vil disse være de eneste som blir brukt på første hopp. Disse inngangsnodene er mottaker av kryptert trafikk som videresendes til mellomnoder.

Mellomnoder

Mellomnoder fungerer som mellomstasjoner, og bidrar til at IP-adressen og identiteten til brukeren skjules. Grunnet informasjon som allerede har blitt kryptert av inngangsnoden, vil det ikke være mulig for mellomnoden å vite hvem brukeren er eller hva slags informasjon som sendes i utgangspunktet. Mellomnoder står ansvarlig for å videresende trafikken til andre noder [34].

Utgangsnoder

Utgangsnoder spiller en sentral rolle i tilgjengeligheten til Tor-nettverkets ressurser. Uten bruken av utgangsnoder ville det ikke vært like lett å tillatte brukere å få tilgang til ressurser på internett. På den andre siden er utgangsnoder med på å svekke anonymiteten på Tor-nettverket. Trafikken som sendes gjennom utgangsnoder kan logges og overvåkes, og dette utgjør en risiko for personvernet. Det kan derfor være en fordel å ikke bruke Tor til å få tilgang til nettsteder som krever pålogging eller inneholder sensitive personopplysninger [35].

2.5.2 Onion-domener

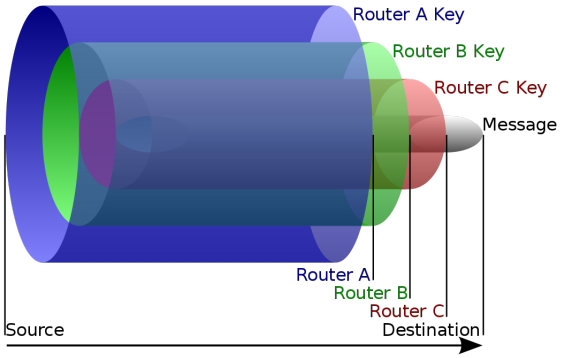
Nettsider og nettjenester som ikke har tilgjengelige IP-adresser, og som ikke kan lokaliseres kalles skjulte tjenester, noe Tor støtter bruken av. URL'en til disse nettsidene har .onion som domene og består av en lang kryptert streng med både tall og bokstaver, som for eksempel:
https://www.nytimesn7cgmftshazwhfgzm37qxb44r64ytbb2dj3x62d2lljsciiyd.onion.
Dette eksempelet viser en V3 lenke, og har en satt lengde på 56 tegn. Den tidligere versjonen V2, hadde bare 16 tegn[36]. Bakgrunnen for å oppdatere til V3 onionlenker er for å øke sikkerheten. Disse nettsidene er kjent for sin anonymitet, dette kommer av at Tor-brukeren og nettsiden blir enige om å møtes på et møtested, på en Tor-node. Ergo verken brukeren eller tjenesten vet hvor motparten befinner seg [37]. Egenskaper som anonym kommunikasjon og deling av informasjon gjør onion-domener attraktive for ulovlig aktivitet. Nettsidene er også kjent for å være svært ustabile, ofte dukker de opp og forsvinner like raskt [38].

2.5.3 Løk-ruting

Teknikken løk-ruting brukes for å anonymisere kommunikasjon på det mørke nettet. Formålet med løk-ruting er å sikre avsenderens identitet overfor mottaker og avlytting underveis. Denne teknikken gjør det vanskeligere å overvåke nettaktivitet og avsløre identiteten til brukere. Betegnelsen løk-ruting kommer av at datapakkene som sendes blir kryptert i flere lag, akkurat som skallene på en løk. Dette er noe som skjer underveis i kommunikasjon for å forsikre at identiteten til avsender forblir anonym. Hvert ledd som håndterer transporten av datapakkene er kun i stand til å dekryptere ett lag, og det benyttes mange lag for å forsikre at verken mottaker eller enkeltledd kan fastslå hele kommunikasjonskanalen [39].

Funksjonen til en løk-ruter er tilnærmet lik en ordinær internett-ruter, og derfor kan all datatrafikk overføres ved hjelp av løkruting. Dette skal være gjennomførbart uavhengig av applikasjonsprotokollen. Den eneste forskjellen på en ordinær internett-ruter og en løk-ruter er at rutingsprotokollene er utformet litt annerledes på nettverkslaget. Data som sendes over et nettverk bestående av løkrutere vil transporteres ved hjelp av en kryptert forbindelse. Denne krypterte for-

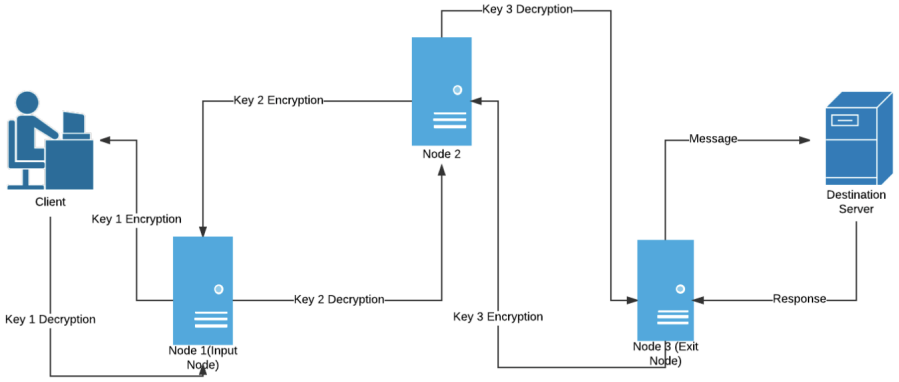
bindelsen settes opp av nettverket mellom en inngangs- og utgangsnode, på denne måten velges det en tilfeldig rute gjennom nettverket av løk-rutere for hver ny tilkobling, og etter en viss tidsperiode [39].



Figur 2.7: Krypteringslag[9]

Datapakker bestående av informasjon som sendes gjennom nettet av løk-rutere vil først krypteres med den offentlige nøkkelen til den neste noden, som heretter refereres til som målnode. Videre vil den krypterte informasjonen, samt adressen til målnoden, bli kryptert nok engang med en offentlig nøkkel. Denne tilhører en ruter med forbindelse direkte til målnoden. Dette er med på å skape en kryptert tunnel som benyttes til å sende informasjon fra en avsender til mottaker. Løk-ruterene på veien fjerner hvert sitt lag med kryptering før datapakken sendes videre til den adressen som blir synliggjort ved dekryptering, som det er mulig å se i figur 2.7. En fare ved å benytte seg av et løk-ruting-nettverk er at din egen maskin kan bli brukt som en node. Årsaken til dette er, som nevnt tidligere, at datapakken har målnoden som avsender hos mottaker. Det kan medføre at en kriminell handling utført av andre kan spores tilbake til din maskin fordi maskinen din står som målnode [40].

[9] https://en.wikipedia.org/wiki/Onion_routing/



Figur 2.8: Løk-ruting krets[10]

Figur 2.8 vist ovenfor gir et bedre innblikk i løk-ruting metoden.

1. Det hele starter med en klient (for eksempel en laptop). Klienten har tilgang til alle krypteringsnøklene, altså nøkkel 1, 2 og 3, og bruker disse til å kryptere GET-forespørselen (meldingen) tre ganger. Dette blir gjort ved å pakke inn meldingen under tre lag som en løk(se figur 2.7), og lagene kan bare skrelles av én om gangen.
2. Denne krypterte meldingen sendes deretter til den første serveren, Node 1.
3. Node 1 har kun adressen til Node 2 og nøkkel 1. Først vil Node 1 prøve å dekryptere meldingen ved hjelp av nøkkelen sin. Den dekrypterte meldingen vil fortsatt være uforståelig da det enda er to lag til med kryptering. Når Node 1 registrerer dette så sendes meldingen videre til Node 2.
4. Node 2 har nøkkel 2 og adressene til Node 1 og Node 3, som også kalles inngangsnode (Node 1) og utgangsnode (Node 3). Etter å ha dekryptert meldingen ved hjelp av nøkkel 2, vil Node 2 registrere at det fortsatt er et lag til med kryptering og videresende meldingen til utgangsnoden.
5. Node 3 (utgangsnoden) skreller av det siste laget med kryptering og finner en GET-forespørsel for nettsiden og sender den videre til destinasjonsserveren.
6. Serveren behandler forespørselen og sender den ønskede nettsiden som svar.
7. Svaret transporteres gjennom de samme nodene i motsatt retning, og hver node legger på et lag med kryptering ved hjelp av sin spesifikke nøkkel.
8. Klienten mottar til slutt et tredobbelt kryptert svar som kan dekrypteres av klienten ved bruk av alle nøklene som klienten har tilgang til [34].

[10] `https://www.geeksforgeeks.org/onion-routing/`

Løk-ruting som teknologi anses som veldig sikker. En er nødt til å knekke flere lag med kryptering, eller kontrollere alle nodene som velges når ruten for overføringen bestemmes for å kunne avdekke identiteten. Likevel kan handlinger utført av en datapakke hos mottaker kompromittere anonymiteten. Det vil si at selv ikke løk-ruting kan bevare anonymiteten til en bruker dersom brukeren spores av informasjonskapsler, utfører handlinger som kan identifiseres eller bruker innloggingstjenester. Tidligere versjoner av Tor og flere andre implementasjoner av løk-ruting avslørte nettsidene som ble besøkt av brukere siden DNS-oppslag ikke ble sendt gjennom løk-ruting-nettverket [39].

2.6 Søkerobot

En søkerobot, også kjent som en crawler eller spider, brukes til å indeksert innhold på nettsider [41]. Søkeroboten er programmert til å systematisk jobbe seg gjennom alt av innhold på nettsidene, inkludert URLer, og samle informasjon som tekstinnhold, metadata og bilder. Eksempler på kjente søkeroboter er Googlebot og Bingbot, som brukes av Google og Bing til å bygge opp databasen til deres søkemotor.

Det finnes flere verktøy som kan brukes til å starte utviklingen av en søkerobot, noen krever programmeringskunnskaper for å benytte, mens andre har et enkelt grafisk brukergrensesnitt som kan tas i bruk [42].

2.6.1 Bibliotek

Søkeroboten som tas i bruk for å løse oppgaven er egenprodusert. I den forbindelse har det blitt tatt i bruk flere biblioteker for å oppnå ønsket funksjonalitet. Her blir de mest sentrale bibliotekene og deres bruksområder nevnt. Bibliotekene som har blitt implementert i sluttproduktet vil bli diskutert under avsnitt 5.4.3.

Beautiful Soup

Beautiful Soup er et Python kodebibliotek som brukes til parsing av data, og er kjent for sin fleksibilitet og enkelhet. Parsing er det å dele opp og strukturere data slik at det kan tolkes og bearbeides av dataprogrammer. Det er en form for maskinell syntaktisk analyse av data, og informasjonen som parses er tekst [43]. Biblioteket ekstraherer data fra HTML- og XML-dokumenter ved å analysere strukturen til en nettside. Beautiful Soup gjør det også mulig å navigere gjennom HTML-strukturen til en nettside og kun hente ut bestemte elementer helt automatisk [44].

Requests

Requests er et bibliotek i Python som gjør det mulig å sende HTTP-forespørsler til en spesifisert URL, dette gjøres ved hjelp av GET-, POST-, PUT-, DELETE- eller

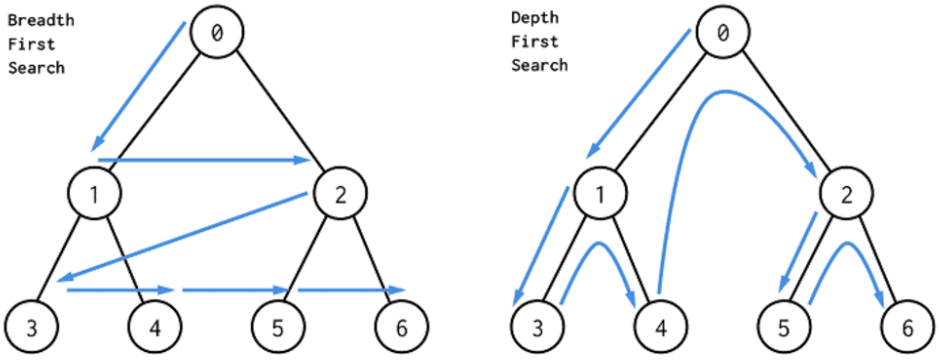
PATCH-forespørsler. Biblioteket er enkelt å bruke og er et av de mest populære innenfor sitt felt. I tillegg til dette har Python Requests kapasitet til å håndtere både forespørsler og responser. Det er også enkelt å legge til parametere, header og andre data i forespørselen, samt å behandle responsen som returneres [45].

Selenium

Selenium er et verktøy som er tilgjengelig på flere forskjellige programmeringsspråk, blant annet Python, og muliggjør automatisering av interaksjon med en webapplikasjon. Den automatiserte interaksjonen blir håndtert som om det skulle blitt utført manuelle handlinger. Selenium har en rekke verktøy og biblioteker for å automatisere funksjoner, som for eksempel åpning av en nettleser, navigasjon inne på nettsiden, utfylling av skjemaer, klikke på lenker og knapper, og hente ut data fra nettsider [46]. For at Selenium skal kunne fungere må det ha en driver, det er denne driveren som starter, kjører og styrer nettleservinduet som blir startet av programmet.

2.6.2 Søkerobotens potensielle traverseringsalgoritmer

Søkeroboten vil trenge en algoritme for hvordan den skal gå gjennom URLene. Alternativene er bredde-først-traversering eller dybde-først-traversering, disse er vist under i figur 2.9.



Figur 2.9: Bredde-først-traversering og Dybde-først-traversering[11]

Figur 2.9 viser traverseringsmetodene i form av søk gjennom trær, men figuren er brukt som et eksempel på å vise hvordan en potensiell traverseringsmetode vil fungere. Den aktuelle URLen er representert ved noden med nummer 0, altså den URLen en starter fra.

[11] https://dev.to/danimal92/difference-between-depth-first-search-and-breadth-first-search-6om

Bredde-først-traversering

Bredde-først-traversering går fra den første URLen som sendes med og lager en kø med URLer. Nye URLer som oppdages legges til bakerst i køen, noe som blir gjort for å sikre at URLen som blir funnet ikke besøkes før resten som ligger i køen har blitt besøkt. Denne prosessen vil gjentas for hver URL som befinner seg på nettsiden. Når det ikke er flere URLer igjen å finne på den gjeldende nettsiden, vil den gjeldende URLen til nettsiden søkeroboten er på, fjernes fra køen. På denne måten vil programmet kjøre en bred traversering, som lar søkeroboten nå ut til et bredere utvalg av nettsteder.

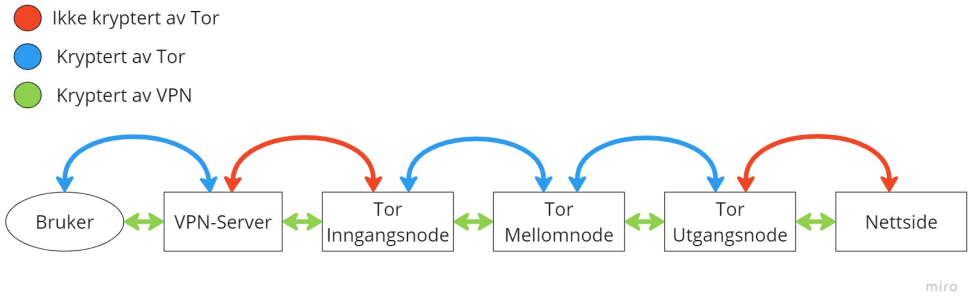
Dybde-først-traversering

Dybde-først-traversering går fra den første URLen som sendes med, og lager en stakk med URLer. I denne stakken vil nye URLer som oppdages, legges til fremst. Slik at URLen som blir funnet sist, er den neste som søkeroboten skal kjøres på. På denne måten vil programmet kjøre en dyp traversering, som lar søkeroboten nå ut til et mindre bredt utvalg av nettsteder. I gjengjeld vil denne traverseringsmetoden bidra til større sannsynlighet for funn av filer eller annen data relatert til nettstedet.

2.6.3 Bruk av VPN til søkeroboten

Som nevnt i avsnitt 2.5.1, tildeler Tor-nettleseren tre noder som skal fungere som de eneste inngangsnodene til en bruker over en lengre tidsperiode. Det er maskinene som kjører søkeroboten som utveksler informasjon med inngangsnodene. En slik utveksling uten bruk av VPN vil inneholde IP-adressen til de respektive maskinene. Dette kan føre til anonymitetsbrudd hvis en aktør overvåker inngangsnoden. Ved bruk av en VPN-tjeneste blir det lagt på et ekstra sikkerhetslag, både i form av krypterte forespørsler og IP-adresser som ikke kan spores tilbake til maskinene [23].

Det er ikke bare inngangsnodene i Tor-nettverket som kan utnyttes av aktører. Som forklart i avsnitt 2.5.1, kan også utgangsnodene brukes til å overvåke trafikken til brukerene. Dersom en VPN er brukt, vil trafikken være kryptert hele veien til nettsiden og tilbake igjen som vist i figur 2.10.



Figur 2.10: Flytdiagram over kryptert trafikk

2.6.4 NC3 - ekspertintervju

Før søkeroboten skulle testes på det mørke nettet ble det arrangert et møte med Nasjonalt cyberkrimsenter (NC3), som er en spesialavdeling under Kripos. NC3 er det nasjonale senteret for forebygging, avdekking og bekjempelse av trusler og kriminalitet over det digitale rom [47]. Møtet ble holdt over kommunikasjonsplattformen Microsoft Teams, og handlet i hovedsak om metoden som blir brukt når NC3 leter etter cyberkriminelle. Her kom det frem at NC3 bruker en del skrapeverktøy, som søkeroboter, for å finne data som kan brukes til etterforskning. Disse vektøyene har de kjøpt tilgang til fra legitime selskap. På det mørke nettet er CAPTCHA ganske utbredt, og dermed må ansatte ved NC3 manuelt sitte å løse disse under skrapeprosessen. For å ikke avsløre for mye angående verktøyene de bruker, ble det ikke sagt noen spesifikk pris for dette, men det kom frem at slike tjenester er veldig kostbare. Fremgangsmåten brukt av NC3 for å finne hackergrupper er veldig mye manuelt arbeid i form av leting gjennom forskjellige forum og kommunikasjonsplattformer. Ved et spørsmål angående bruk av proxy-tjenester mot VPN ble det svart at det kom opp til hvem en kan stole på. Objektivt sett vil både proxy og VPN hjelpe med å anonymisere bruker, men VPN er det beste valget av disse to. Se vedlegg G for utfyllende møtereferat.

Kapittel 3

Metode

3.1 Introduksjon

Dette kapitlet er delt inn i flere deler basert på de ulike metodene som ble brukt for å gjennomføre oppgaven i sin helhet. Under hver del blir det beskrevet hvilken metode som har blitt brukt og begrunnelse for avgjørelsene som har blitt tatt. Først og fremst beskrives metodene som ble brukt for datainnsamling, deretter utdypelsen rundt metodene brukt gjennom utviklingen av søkeroboten, samt oppbyggingen av denne.

3.2 Datainnsamling

Det ble bestemt datainnsamlingsmetode på bakgrunn av problemstillingen og problemformuleringen. Som beskrevet der vil det i denne oppgaven være flere ulike metoder knyttet til å nå målene med oppgaven. For å kunne oppnå disse var en altså nødt til å benytte en kombinasjon av kvantitativ og kvalitativ metode. I løpet av oppgaven ble det blant annet gjennomført både en litteraturstudie og et utviklingsarbeid. I tillegg ble det utført en spørreundersøkelse blant gruppens nettverk.

Behovet for dybde i oppgaven krevde en kvalitativ tilnærming slik at det var mulig å få en bred forståelse for feltet, herav informasjonsinnhenting om Tor, bøker og artikler som omhandler forskning og de positive aspektene ved Tor. Det var likevel nødvendig å ha data som kunne tallfestes, derav kvantitativ tilnærming. Ønsket fra oppdragsgiver var å høste en mengde med data fra Tor som videre kunne tallfestes, analyseres og forskes på i egen doktorgradavhandling.

3.3 Litteraturstudie

En litteraturstudie er en metode som brukes for å oppsummere eksisterende litteratur og eventuell forskning på et problemområde, og det skal gjøres på en systematisk måte. En gjennomgår tidligere litteratur i feltet og analyserer disse, som

videre kan føre til innsikt og andre perspektiver, for eksempel for å svare på problemstillinger, men det medfører ikke direkte ny kunnskap [48].

For denne oppgaven ble det brukt litteraturstudie for innhenting av informasjon som legger til grunn for det teoretiske fundamentet for oppgaven. For å få tak i informasjon om bakgrunn, historie og teknisk oppbygging av Tor, var det bruk av internettkilder i hovedsak. Tor Prosjektets selvpålagte oppdrag vekket gruppens nysgjerrighet og det ble bestemt å utforske dette området. Dette var også i samråd med oppdragsgiver, som mente det var ønskelig at det ble fokusert på den fornuftige bruken av Tor. På denne måten ble det valgt å gjøre systematiske søk, og analyser av disse søkene.

3.3.1 Bakgrunn, historie og teknisk oppbygging

Dette avsnittet forklarer fremgangsmåten for å skrive om bakgrunn, historie og teknisk oppbygging. For å skape en bred forståelse og innsikt i Tor ble informasjonshentingen her utført, som nevnt, ved å bruke internettkilder. Det ble ikke utført systematiske søk eller gransking av litteratur på disse delene av litteraturstudien. De ulike delene av internettet, det åpne, det dype og det mørke nettet er med for å vise at internett består av ulike nivåer, og slik at leseren skulle forstå forskjellen. Fakta ble samlet inn og isfjell-analogien ble brukt, en ganske vanlig analogi for å forklare noe som er synlig og noe som er usynlig (se figur 2.1). Dette valget ble tatt blant annet for å belyse at for å nå det mørke nettet må det gjøres med viten og vilje.

For å få en bred forståelse av Tor på et dypt nivå, så gruppen det som absolutt nødvendig å ha med bakgrunn og historie, hele Tors opprinnelse. Det ble satt like krav til kildene som ble brukt for denne informasjonshentingen som inklusjonskriteriene nevnt nedenfor under avsnitt 3.3.3. Bakgrunnen og historien til Tor henger litt sammen og det ble bestemt å forsøke å skrive det i en kronologisk rekkefølge, slik at det skulle bli mest mulig ryddig og lettere for leseren å henge med. Det ble utelukkende valgt å kun inkludere historiske hendelser som hadde en sammenheng med Tors opphav og utvikling å gjøre.

En naturlig del av å forstå funksjonen til et verktøy er å tilegne seg kunnskap om den underliggende teknologien det er basert på. Kapitlet om Tors tekniske oppbygging er basert på informasjonsinnsamling for å oppnå en dypere forståelse av teknologiens virkemåte, samt for å legge grunnlaget for utviklingen av en optimal søkerobot på det mørke nettet. Underkapitlene omfatter Tor-nettverket, løkruting, kryptering og onion-domener, som anses som nødvendig informasjon for en helhetlig forståelse av oppgaven. Oppdragsgiver, på grunn av egen off. PhD, har vært kjent med flere pålitelige nettressurser og har bidratt aktivt til å finne troverdige kilder, som har blitt hyppig brukt i dette kapitlet. I tillegg ble det utført søk i Google for å finne relevante kilder. Både norske og engelske søkeord

ble benyttet, og det ble observert at engelskspråklige søk ga flere omfattende og informative resultater. Søkeordene "The onion router", "onion routing", "onion routing protocol", "løkruting" og "Tor network" ble benyttet. Alle søkeordene resulterte i relevante kilder, men et utvalg ble foretatt i tillegg til kildene som ble levert av oppdragsgiver. Boken *Digital Sikkerhet: en innføring* ble også benyttet, da gruppen allerede hadde kjennskap til denne boken gjennom studieløpet. Blant søkeresultatene var det også mulig å finne forskningsartikler, men disse artiklene fokuserte primært på forfatternes eller forskernes egen forskning knyttet til Tor, i stedet for den underliggende teknologien. Derfor ble det ansett som nødvendig å finne informative nettressurser som kunne forklare teknologien på en forståelig måte, da medlemmenes kunnskapsgrunnlag om Tor var begrenset.

3.3.2 Fordeler med Tor

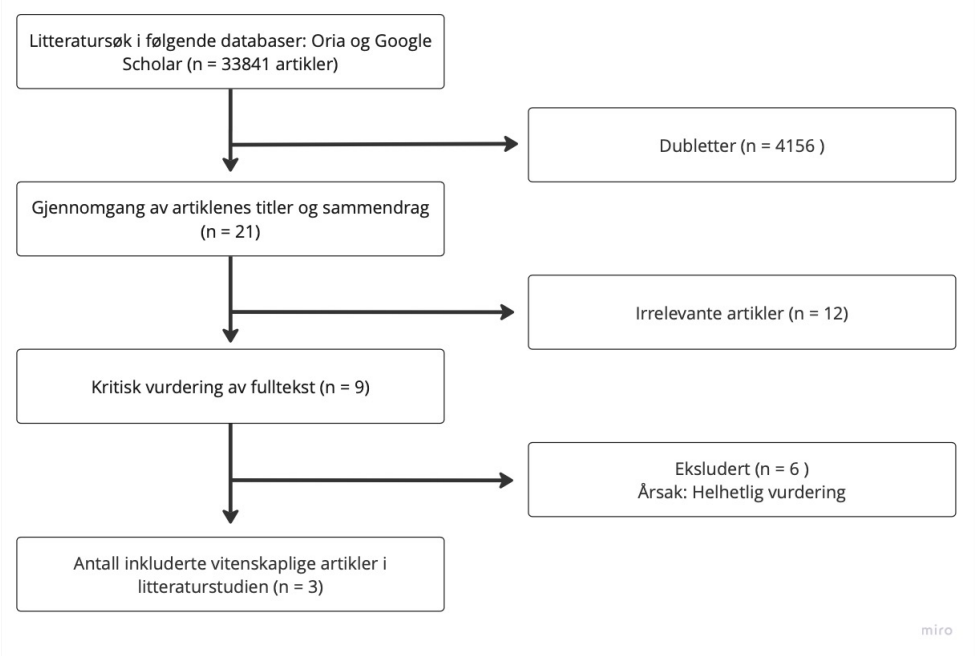
Dette avsnittet beskriver hvordan litteraturstudie for utforskning av fordeler foregikk. For å finne kilder til litteraturen ble det bestemt å bruke *Google Scholar* og *Oria* som databaser å lete i. *Google Scholar* er Google sin søkemotor som kun inneholder akademisk litteratur og *Oria* er NTNUs universitetsbibliotek på nett. Som en kan se i oversikten under avsnitt 3.3.3, ble det brukt noen søkeord for å forsøke å få frem relevant litteratur. Søkene ble utført med engelske søkeord ettersom dette ga flere søkeresultater som omfattet ønsket tema. Det finnes en del forskning på det mørke nettet og Tor, men det er forholdsvis lite sammenlignet med andre områder. I tillegg var det viktig at litteraturen vi fant, som omhandlet Tor, også analyserte eller diskuterte fordeler med Tor. Selv om databasene som ble brukt ville vise litteratur med god kvalitet, måtte hvert funn analyseres for relevans og kvalitet.

3.3.3 Inklusjonskriterier litteraturstudie

Dette avsnittet er med for å vise kravene som er satt for å velge litteratur.
Det er viktig at informasjon og data som brukes er troverdig. Derfor ble det lagt vekt på hvem forfatteren var, om det var en privatperson eller en organisasjon og utgiver. Listen for kriterier kan sees under her.

- Litteratur som ble brukt måtte være på et skandinavisk språk eller engelsk
- Aktuelle hovedsøkeord:
    1. The Onion Router
    2. Tor Browser
    3. The dark web
- Årstall, fra 2010 og frem til nå
- Litteraturen måtte omhandle problemstillingen

Nedenfor vises et flytdiagram som illustrerer antallet treff i de valgte databasene, samt antall artikler som ble samlet inn og vurdert basert på tittel og sammendrag. Videre presenteres også artiklene som ble kritisk vurdert og ekskludert etter en kvalitetsvurdering. Til slutt blir artiklene som ble inkludert, presentert.



Figur 3.1: Artikkelflyt for litteraturstudien

3.3.4 Søkestrategi

I følgende søkestrategitabell presenteres en oversikt over brukte søkeord, antall treff generert av hvert søkeord og valgte database. Kolonnen "aktuelle funn" dokumenterer artikler som ble inkludert i den kritiske vurderingen.

Søkemotor / database	**Søkeord**	**Aktuelle funn**	**Antall treff**
Oria.no	The Onion Router	"Mapping Real-World Use of the Onion Router" "Dilemmas related to the functioning and growth of Darknet and the Onion Router network"	279
Oria.no	Tor browser	"The Tor Browser and Intellectual Freedom in the Digital Age" "Tor, what is it good for? Political repression and the use of online anonymity-granting technologies" **"The potential harms of the Tor anonymity network cluster disproportionately in free countries"**	805
Oria.no	"Tor" AND "The dark web"	**"Hands-On Dark Web Analysis"**	107

Google scholar	The Onion Router	**"The onion router: Understanding a privacy enhancing technology community"**	12600
Google scholar	Tor browser	"New Me: Understanding Expert and Non-Expert Perceptions and Usage of the Tor Anonymity Network"	14100
Google scholar	"Tor" AND "The dark web"	"The Dark Web Dilemma: Tor, Anonymity and Online Policing"	5950

Tabell 3.1: Søkestrategi

Som vist i tabell 3.1 ble det også gjennomført et søk med søkeordene "Tor" og "dark web" kombinert for å finne forskningartikler som omhandlet det mørke nettet og Tor. Tabellen dokumenterer alle forskningartiklene som det har blitt tatt en kritisk vurdering av, det vil si at artikkelen har blitt gjennomgått i full tekst, og det ble tatt en vurdering på om litteraturen inneholdt det gruppen ønsket å fremme. Grunnet omfanget av litteraturstudien var det kun et begrenset antall med artikler som ble inkludert, og derfor er det ikke mulig å si at litteraturen som ble ekskludert ikke var relevant for problemstillingen. Etter gjennomgangen av all litteratur ble det konkludert med at alle artiklene kunne vært brukt, men det hadde allerede blitt utført informasjonsinnhenting ved hjelp av internettkilder for å tilegne grunnleggende kunnskap om bakgrunn og teknologien, og derfor ble det sett på som overflødig å benytte forskningartikler til å validere informasjonen som hadde blitt funnet på blant annet Tor Prosjektets opprinnelige nettside. Det ble valgt ut tre forskningsartikler som ble inkludert i rapporten. Disse artiklene har blitt fremhevet i tabell 3.1. "Hands-on Dark Web Analysis" ble utgitt i 2018 og er skrevet av Sion Retzki, en IT-ekspert med 20-års erfaring innenfor fagfeltet [14]. "The onion router: understanding a privacy enhancing technology community" er en artikkel skrevet av Hsiao-Ying Huang og Masooda Bashir utgitt i 2016. Huang har en doktorgrad i informatikk [49], og Masooda Bashir har grader innenfor matematikk, datavitenskap samt en doktorgraden i psykologi [50]. Den siste artikkelen med tittelen "The potential harms of the Tor anonymity network cluster disproportionately in free countries" er skrevet av Eric Jardine, Andrew M. Lindner og Gareth Owenson. Eric Jardine er en cyberkrimforskningsleder hos Chainalysis og har en doktorgrad i internasjonale relasjoner [51]. Andrew M. Lindner er en professor på

Skidmore College, og underviser i emner som kvantitativ undersøkelsesmetode, politisk sosiologi og massemedier [52]. Gareth Owenson er en så kalt Tor ekspert, han har en doktorgrad i datavitenskap [53], og er per i dag CTO og medgründer i Searchlight Cyber, som er en bedrift som tilbyr etterforskningverkøty som kan brukes på det mørke nettet for å avdekke aktivitet av tvilsom karakter [54]. Artiklene som har blitt valgt ut er skrevet av forfattere med god kunnskap innenfor feltet. Resultatene fra denne delen av litteraturstudien har blitt integrert i teorikapittelet av rapporten i samråd med veileder. Videre har sistnevnte artikkel blitt brukt til å definere avsnitt 1.2.

3.4 Spørreundersøkelse

Den bakenforliggende årsaken til utførelsen av spørreundersøkelsen var å få bedre innsikt i kunnskapen folk besitter om det mørke nettet og deres oppfattelse. Gjennom egne erfaringer og samtaler med oppdragsgiver ble det avdekket at det mørke nettet har fått et dårlig omdømme til tross for dets positive opphav. På grunn av spørreundersøkelsens formål, inneholdt den spørsmål som skulle gjenspeile generell kunnskap og oppfatning av det mørke nettet, samt hvordan deltakerene har benyttet seg av teknologien. Undersøkelsen ble utrettet gjennom plattformen Google Forms [55], en plattform som gruppen hadde kjennskap til fra tidligere.

3.4.1 Spørsmålsguide

Seksjoner	Spørsmål
Bakgrunn	Hva er din tekniske bakgrunn?
	Har du noen gang brukt Dark Web før?
	Hva er din kjennskap til Dark Web?
Bruk av det mørke nettet	Hvilke forhåndsregler tar du når du bruker Dark Web?
	Hvilken nettleser bruker du på Dark Web?
	Hvorfor bruker du Dark Web?
Oppfatning	Hva er din oppfatning av lovligheten til aktivitetene som foregår på Dark Web?
	Hvilken informasjon ville du vært villig til å dele om deg selv på Dark Web?

Tabell 3.2: Spørsmål til spørreundersøkelsen

I tabell 3.2 presenteres spørsmålene som har blitt stilt i spørreundersøkelsen, i tillegg til hvilken seksjon av spørreundersøkelsen de tilhører. Spørsmålene som er under seksjonen "Bakgrunn" er de første som ble spurt i undersøkelsen. Neste seksjon, "Bruk av det mørke nettet", inneholder spørsmål som deltakeren kun får mulighet til å svare på hvis de svarte "Ja" på spørsmålet "Har du noen gang brukt Dark Web før? ". Disse spørsmålene går ut på hvordan deltakeren bruker det mørke nettet. Den siste seksjonen, "Oppfatning", inneholder de siste spørsmålene som deltakeren blir stilt. Disse spørsmålene blir ikke påvirket av tidligere svar og vises derfor hos alle deltakere. Bruken av betegnelsen "dark web" var et aktivt valg ettersom dette er den mest kjente og brukte betegnlese både i media og fagfeltet.

Basert på tidligere erfaringer er det en tydelig sammenheng mellom spørreundersøkelsens struktur og svarfrekvens. Derfor er alle spørsmålene som stilles i spørreundersøkelsen blitt presentert med svaralternativer. Hensikten med svaralternativene er å gjøre undersøkelsen så enkel som mulig, derav mer attraktiv å svare på. Høy svarfrekvens tilsvarer større mengde data som kan analyseres og brukes til å besvare oppgaven. Utvalget av spørsmål i undersøkelsen, inkludert spørsmål 1, 2, 3, 5 og 7 som vist i tabell 3.2, ble formulert som et flervalgsspørs-

mål med enkeltsvar. Dette innebar at det ble vist 2-6 svaralternativer under hvert av disse spørsmålene og deltakeren fikk kun velge ett svar. Spørsmål 4 og 8 kunne også defineres som flervalgsspørsmål, men her var det mulig å velge flere svaralternativer. En vanlig utfordring med slike flervalgsspørsmål er begrensningen i svaralternativer som kan føre til partiske resultater. Dersom ingen av de forhåndsdefinerte svaralternativene passer for respondentene, kan dette føre til tilfeldige svarvalg som igjen kan påvirke nøyaktigheten av resultatene. Med dette tatt i betraktning ble spørsmål 6 presentert med en tekstboks, i tillegg til forhåndsdefinerte alternativer for å gi deltakerne friheten til å svare på spørsmålet med egne ord.

3.4.2 Distribusjon av spørreundersøkelsen

Spørreundersøkelsen ble sendt ut via de sosiale kanalene: Discord, Facebook og Messenger. Siden målgruppen til undersøkelsen var nettverket til gruppen, var det disse kanalene som virket mest relevant for best mottakelse. Gjennom Discord kunne undersøkelsen nå studenter som går IT-studier. Gjennom Facebook og Messenger kunne undersøkelsen bli sendt ut til en mer generell befolkning, både med og uten IT-bakgrunn.

3.4.3 Resultater sammenlignet med spørreundersøkelse utført av CIGI

Etter en gjennomgang av resultatene fra spørreundersøkelsen, konkluderte gruppen med at tallene fra egen undersøkelse ikke ga tilstrekkelig representativitet for verdensbefolkningen, og dermed er det ikke mulig å trekke pålitelige konklusjoner. Dette skyldes hovedsakelig begrensningene i språk og media som ble brukt til å dele spørreundersøkelsen, som har resultert i lite mangfold. For å øke troverdigheten til resultatene fra spørreundersøkelsen som ble utført av gruppen, ble det foretatt søk for å finne tilleggsdata som kan kvantifiseres og kobles til egne resultater. Etter en betydelig mengde søk ble spørreundersøkelsene til *Centre for International Governance Innovation* funnet på organisasjonens hjemmeside. Resultatene som har blitt inkludert i denne rapporten stammer fra spørsmål som er tilnærmet lik de som har blitt stilt i gruppens spørreundersøkelse. Disse spørsmålene omhandler oppfatning av det mørke nettet, antall brukere av Tor og årsak til bruk og eventuelt avstå bruk av Tor og/eller det mørke nettet.

3.5 Søkeroboten

I denne delen vil det presenteres avgjørelser i forbindelse med utviklingen av søkeroboten. Før utviklingen av søkeroboten kunne begynne, måtte forutsetninger som programmeringsspråk, utviklingsplattform og generell forståelse av Tor være på plass. Valget av programmeringsspråk og utviklingsplattform er viktige slik at

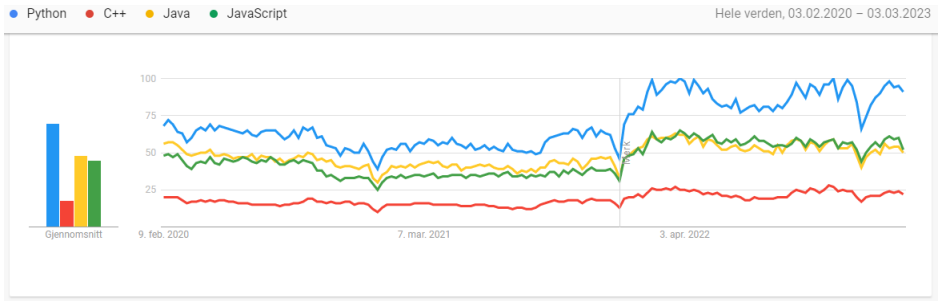
utviklingsprosessen går så problemfritt som mulig. Grunnlaget for å oppnå en forståelse av Tor er at søkeroboten skal kunne traversere på det mørke nettet som er tilgjengelig gjennom Tor-nettleseren. En metode for traversering av nettsider er også et valg som er blitt tatt.

3.5.1 Valg av programmeringsspråk og plattform

Programmeringsspråk

Valget av programmeringsspråk ble gjort ut fra flere kriterier. Det ble sett etter ulike relevante bibliotek, utviklersamfunn og kompatibilitet mellom ulike operativsystemer. Det finnes flere programmeringsspråk som har et bredt utvalg av biblioteker tilpasset for nettskraping. Ved bruk av Python kan en benytte Python Package Index (PyPI) [56], som er et oppslagsverk for bibliotekene som Python støtter.

I figur 3.2 illustreres de mest populære programmeringsspråkene, via Google trends. Her er det også mulig å se hvordan søkene mot de fire mest populære programmeringsspråkene har utviklet seg.



Figur 3.2: Statistikk over aktuelle programmeringsspråk[1]

Gruppen hadde i ulik grad noe kunnskap innen alle disse fire programmeringsspråkene som er vist i figur 3.2. Figuren viser at det har vært en merkbar økning søk på Google, relatert til Python den siste tiden. Når flere engasjerer seg og stiller spørsmål om problemer de støter på ved bruk av et programmeringsspråk, vil mengden ressurser som kan brukes i ettertid øke. Disse ressursene er med på å skape et godt utviklersamfunn. Et godt utviklersamfunn vil si at det er rikelig med dokumentasjon, veiledninger, forum og bidragsytere som deler bibliotek og kodesnutter.

Den største faktoren bak valg av språk var hvor egnet det er til å lage en søkerobot. Det som gjør Python godt egnet til utvikling av søkeroboter er alle de

[1] `https://trends.google.com/trends/explore?date=2020-03-02%202023-03-03&q=%2Fm%2F05z1_,%2Fm%2F07sbkfb,%2Fm%2F0jgqg,%2Fm%2F02p97`

omfattende bibliotekene som er designet for å gjøre automatisering av datahøsting lettere. Tilnærmingen til Python "lite kode, mye funksjonalitet", gjør at koden er lett å implementere og vedlikeholde [57]. Da dette er et programmeringsspråk som gruppen behersker godt, ble det ikke nødvendig å legge mye tid i å lære et nytt språk.

Utviklingsplattform

Hovedtanken for valg av plattform var å bruke Visual Studio Code (VSC) for å utvikle søkeroboten. Dette gir muligheten for å samskrive gjennom utvidelsen Live Share. Det er en utvidelse som lar en dele sin instans av VSC. En av ulempene ved bruken av VSC er at det opprinnelig ikke støtter noen programmeringsspråk. Dette vil si at en må laste ned alle nødvendige utvidelser, og alle biblioteker som er brukt gjennom koden, da disse er lagret lokalt på hver enkelt sin maskin. En av de positive sidene ved VSC er at det er mye brukt i utviklersamfunnet, Microsoft har i tillegg en utdypende dokumentasjonsside som bidrar til mengden ressurser rundt bruken av VSC [58].

Et annet alternativ som ble vurdert var PyCharm [59], da dette er et utviklingsmiljø spesialisert mot programmering i Python. Selv med ulempen til VSC og potensialet av PyCharm, ble VSC regnet som det beste alternativet da det både har et godt utviklersamfunn og er konfigurerbart slik at en kan tilpasse plattformen slik en selv ønsker. Samtidig som VSC er det utviklingsmiljøet som gruppen kjenner best fra før.

3.5.2 Oppsett av PCer

Som nevnt i avsnitt 1.10.2 ble det utdelt fire fabrikkinnstilte Lenovo PCer som skulle brukes som eneste inngangspunkt til Tor-nettverket. Disse maskinene var utdaterte og hadde med dette lav ytelse sammenlignet med nyere enheter. PCene måtte derfor ha et operativsystem som kunne utnytte maskinvaren, og den lave ytelsen på best mulig måte. Oppdragsiver ga en anbefaling om å bruke Linux Lite. Linux lite er et Ubuntu-basert operativsystem med lave krav til maskinvare [60].

Installeringen av Linux Lite på PCene ble gjort gjennom en minnepinne, der operativsystemet var nedlastet. Stegene i Linux Lite sin hjelpemanual [61] beskriver hvordan denne prosessen gjennomføres og det var denne gruppemedlemmene brukte for å få Linux Lite på disse PCene. Rufus, som er et program for å omformatere en minnepinne om til en oppstartbar minnepinne [62], ble brukt for å gjennomføre denne prosessen.

Dataene som lagres og legges over på harddisken, som nevnt i avsnitt 1.10.2, må kunne analyseres både ved bruk av Windows og Linux. For å få til dette er det viktig at filsystemet på harddisken er kompatibel med de nevnte operativsystemene. Et av filsystemene som brukes på tvers av operativsystem er Extensible File

Allocation Table (exFAT) [63]. Da exFAT allerede var filsystemet til harddisken gruppen hadde mottatt av oppdragsgiver, falt valget på å beholde dette filsystemet. I tillegg er exFAT det filsystemet som er mest kompatibelt med hvilket som helst system [63].

3.5.3 Valg av traverseringsalgoritme

Valget av traverseringsalgoritme falt på bredde-først, da denne algoritmen lar søkeroboten nå ut til et bredere utvalg av nettsteder. Det er vanskelig å si noe om hvor dypt det er mulig å komme på ulike nettsteder. Dette betyr at dybde-først-traversering kan følge en URL i veldig lang tid før den returnerer tilbake til start-URLen. For at søkeroboten skulle fange opp mest mulig spredt informasjon, ble bredde-først valgt som traverseringsalgoritme.

Oppsett av VPN og valg av VPN-tjeneste

Koblingen ble satt opp på en Raspberry Pi som hadde installert programvaren RaspAP. Det er en programvare som fungerer for flere ulike Linux operativsystemer, og gjør enheten om til en trådløs ruter. RaspAP ble valgt grunnet det enkle og brukervennlige oppsettet sammen med at det var relativt enkelt å gjøre konfigurasjoner. Den viktigste konfigureringen som ble gjort var installeringen av OpenVPN. OpenVPN er et åpent kildekode prosjekt, med en egen protokoll "OpenVPN tunneling protocol". VPN som ble satt opp til bruk av søkeroboten kan endre mellom å ha IP-adresse i Danmark og i USA. USA ble valgt som den ene lokalisasjonen grunnet store mengder brukere i landet. Tilkoblingen til Danmark ble satt opp slik at ruteren fort kunne konfigureres om for kortere hopp, som kan føre til økt nettverkshastighet.

Et av gruppemedlemene hadde tilgang til VPN-tjenesten ExpressVPN, som er kjent for å være både rask og sikker [64]. Den eneste informasjonen som blir lagret av ExpressVPN er dato ved tilkobling, anonymisert data om ExpressVPN-applikasjonen som blir brukt, i tillegg VPN-server lokalisasjon koblet opp til og total mengde med data som blir overført hver dag [65]. Ingen av disse dataene kan spores tilbake til brukeren, verken av andre eller av ExpressVPN. Grunnet alle de positive sidene ved ExpressVPN, ble det valgt å bruke denne.

3.6 Utvikling og oppbygging av søkeroboten

Gruppen valgte å ta på seg oppdraget med å utvikle en egen søkerobot, istedenfor å finne en ferdigutviklet søkerobot. Dette både for å få testet og videreutviklet programmeringskunnskapene, men også fordi det var en utfordring gruppen ville forsøke seg på. Ved å utvikle en egen søkerobot krever det dypere forståelse av hvordan Tor-teknologien er bygd opp, i tillegg vil en ha full kontroll på hvordan data hentes ned og hvor hyppig nettverksforespørsler sendes. Det ble avdekket i

møtet med NC3 (vedlegg G), at ferdigutviklede søkeroboter, som er av god kvalitet og kan håndtere det mørke nettet, ofte er kostbare.

Søkeroboten er bygd opp med to hovedfunksjonaliteter. Disse er funksjonaliteter som gjør det mulig å skrape en nettside, og å krype videre fra en nettside til en annen. Skraping vil si at søkeroboten høster data i form av tekst, bilder eller filer fra en nettside [66]. Kryping vil si at søkeroboten kan forflytte seg videre til de lenkene som er referert til på den aktuelle nettsiden [67].

3.6.1 Programmets startargumenter

For å kunne kjøre søkeroboten må den ha en tekstfil med en eller flere URLer, og et forhåndsdefinert valg om søkeroboten kun skal skrape, eller både krype og skrape. Denne tekstfilen vil være nødvendig for at programmet skal kunne kjøre. Som vist i kodeliste 3.1 vil filen med URLer legges til hvis en bruker kommandolinjeargumentet *-i* med parameteren *inputfile*, som er navnet på tekstfilen.

Kodeliste 3.1: Argumenter til bruk ved start av søkeroboten

```
"""
    scraper.py -<argument>

    Arguments:

    -i <inputfile> "File containing urls to crawl/scrape from"
    -s "Adds only scraping functionality to the program"
    -c "Adds crawling and scraping functionality to the program"

"""
```

For at programmet skal vite om det kun skal skrape, eller krype og skrape, må det sendes med et kommandolinjeargument. Enten *-s* eller *-c*. Som vist over i kodeliste 3.1 vil *-s* gi funksjonalitet for kun skraping på URL-listen som er sendt med, *-c* vil gi funksjonalitet for både kryping og skraping.

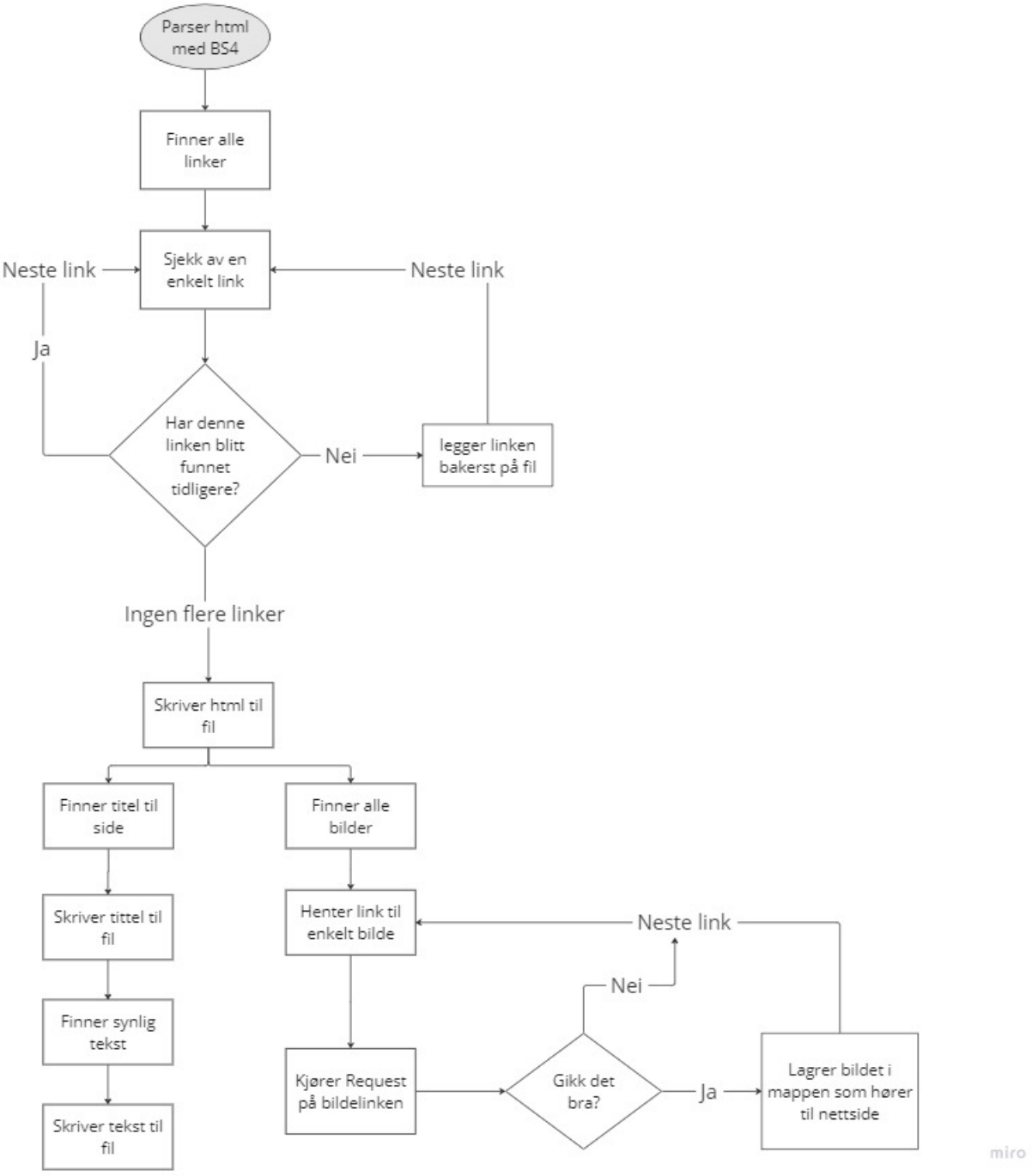
3.6.2 Søkerobotens hovedkomponenter

Første versjon av søkeroboten benytter seg av bibliotekene Requests og Beautiful Soup. Programflyten kan sees i vedlegg D. Den siste versjonen av søkeroboten benytter seg av bibliotekene Selenium og Beautiful Soup, og programflyten på det kan sees i vedlegg E.

Første versjon - Requests og Beautiful Soup

Første versjon av søkeroboten ble bygget rundt biblioteket Requests. Det var flere grunner til valget om bruk av Requests, blant annet egne erfaringer med biblioteket fra tidligere emner. Hovedgrunnen til dette valget var at Requests et bibliotek som er kjent for raske nettverksforespørsler, både i form av sending og mottak av

data. Det fungerte veldig bra på det åpne nettet og nettsteder på det mørke nettet som kjørte over https. Derimot når den første versjonen av søkeroboten skulle testes på nettsider som kjørte over http, endte det i flere komplikasjoner. Søkeroboten ville ikke sende forespørsler til sidene, og det kom opp flere feilmeldinger. Feilmeldingene var lite informative, og var derfor vanskelig å fikse. Beautiful Soup ble brukt til selve uthentingen av data. Data i form av URLer, HTML, tekst, tittel og bilder ble forsøkt hentet fra hver enkelt nettside, som vist i figur 3.3.



Figur 3.3: Programflyt for hvordan Beautiful Soup fungerer i første versjon

Siste versjon - Selenium og Beautiful Soup

I den siste versjonen av søkeroboten er Requests byttet ut med Selenium. Selenium ble valgt for at søkeroboten skulle kunne etterligne menneskelig adferd. På det mørke nettet er det veldig mange nettsteder som benytter seg av CAPTCHA. Det ble oppdaget at dersom nettstedene mistenkte at det var en robot som sendte forespørselen, forekom CAPTCHA hyppigere. Søkeroboten ble programmert til å vente før den begynte å skrape både tekst og bilder. Beautiful Soup er fortsatt brukt, men i litt mindre grad enn i den første versjonen. I den siste versjonen ble Beautiful Soup kun brukt for å høste data i form av HTML, synlig tekst og tittel på nettsiden.

3.6.3 Programmets operasjoner/livsløp

Kildekoden til den seneste versjonen av søkeroboten ligger i vedlegg F. Flytdiagram av den aktuelle versjonen av søkeroboten ligger i vedlegg E.

Som nevnt i avsnitt 3.6.1 måtte søkeroboten få tilsendt startargumenter før den kunne starte å kjøre. Programmet startet så en Selenium-driver. Denne driveren fungerte som en kontroller for søkeroboten. Det første denne driveren gjorde var å koble Tor-nettleseren til Tor-nettverket. Programmet ventet så i en tilfeldig tid mellom 33 og 43 sekunder for å holde roboten så lik et menneske som mulig, samtidig som dette ga søkeroboten tid til å koble seg til Tor-nettverket. Ved tidligere observasjoner ble søkeroboten for rask, og stoppet før den kunne koble seg til nettverket.

Krypefunksjonalitet

Som vist i både kildekoden (vedlegg F) og flytdiagrammet (vedlegg E), sørget krypefunksjonaliteten først for å opprette hovedmappene til databasen. Deretter sjekket programmet om domenet har blitt besøkt tidligere. Gjennom testingen av søkeroboten ble det oppdaget at dersom en nettside brukte over 30 sekunder på å svare, førte det til et tidsavbrudd grunnet Selenium sine standard innstillinger. Tor-nettverket sin kompleksitet gjør det til en utfordring å opprettholde en rask og effektiv nettverkstilkobling, som ofte resulterer i at nettsider bruker lengre tid på å laste inn. Derfor ble det valgt å øke ventetiden før et tidsavbrudd til 60 sekunder. I tillegg til en økning av ventetid, ble det lagt til at hver side som ikke svarte skulle testes tre ganger før den URLen ble ignorert. Dette ble gjort for å forsikre at ingen nettsider ble oversett. Dersom domenet var tilgjengelig og ble tilkoblet, kalte programmet skrapefunksjonaliteten, se neste avsnitt *skrapefunksjonalitet*. Etter at dataen ble høstet fra domenet, hentet programmet ned alle aktuelle URLer som var tilgjengelige på nettsiden, og plasserte disse i en liste. Programmet prøvde så å hente ned neste URL på listen, og denne prosessen gjentok seg til alle URLer i listen var gjennomgått.

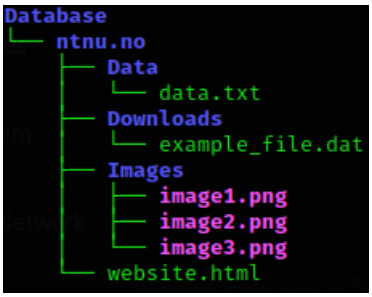
Skrapefunksjonalitet

Som vist i både kildekoden (vedlegg F) og flytdiagrammet (vedlegg E), sørget skrapefunksjonaliteten først for å sjekke om data allerede hadde blitt høstet fra domenet. Dersom det ikke var tilfellet ble undermappene til det aktuelle domenet opprettet. Programmet parset så HTMLen som nettsiden hadde. Dette ble gjort ved bruk av Beautiful Soup. Etter at HTMLen ble lagret, høstet programmet ned tittelen på nettsiden og tekst som var synlig.

Før programmet startet med å høste bilder, ventet programmet i 12 sekunder. Dette ble gjort for at alle tilgjengelige bilder skulle få tid til å laste inn. Det ble også ventet 2 sekunder mellom hvert bilde som ble høstet, da enkelte nettsider venter med å laste inn bilder til brukeren blar ned på siden. Grunnen til at bildene måtte bli lastet inn, var som følge av funksjonen "screenshot", som tar et skjermbilde rundt det objektet som ble sendt med [68]. Etter dette så programmet etter knapper som inneholdt ordet: *Download*. Dette ble gjort for at programmet skulle vite hvilke knapper som måtte trykkes på for å kunne laste ned filer som var tilgjengelige på nettsidene.

3.6.4 Database til høstet data

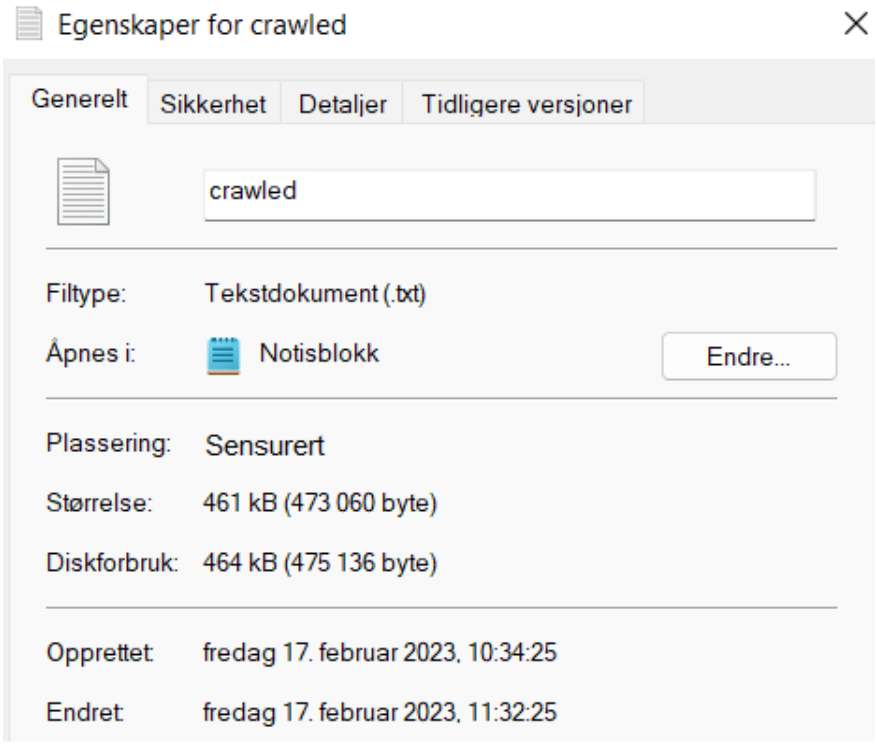
Søkeroboten høstet data fra nettsider den besøkte. Dataene som ble høstet ble plassert i en egen mappe, som kunne gjenkjennes på navnet til nettsiden, altså URLen. Dette var da hovedmappen i databasen til akkurat dette nettstedet. Under denne mappen ble det laget tre nye mapper med navnene "images", "data" og "downloads". I mappen "images" ble alle bildene som var på nettsiden lagret. I mappen "data" ble all synlig tekst på nettsiden lagret, og i "downloads" ble alle nedlastede filer fra nettsiden plassert. I tillegg til synlig tekst, bilder og filer, ble en HTML-fil av nettsiden laget og lagret i hovedmappen til nettsiden. Figur 3.4 gir et eksempel på hvordan databasestrukturen så ut. Figuren ble laget med utgangspunkt i høsting av data fra nettsiden ntnu.no.



Figur 3.4: Eksempel på databasestruktur

3.7 Testing av søkeroboten

Testing ble gjort delvis parallelt med utviklingen for å teste at alle funksjoner fungerte slik som ønsket. I startfasen ble all testing gjort på det åpne nettet mot ulike nyhetssider. Her kunne søkeroboten enkelt testes om den kunne krype gjennom nettsider og skrape med seg både tekst og bilder. Figur 3.5 viser hvordan søkeroboten fant frem til 461 kB med kun URLer, som tilsvarer 5492 unike URLer. Testkjøringen varte i 58 minutter, som kan sees i feltene *Opprettet* og *Endret* på figuren. Feltet *Plassering* er blitt sensurert, da testen ble gjennomført på en personlig PC.



Figur 3.5: Resultat av test på det åpne nettet

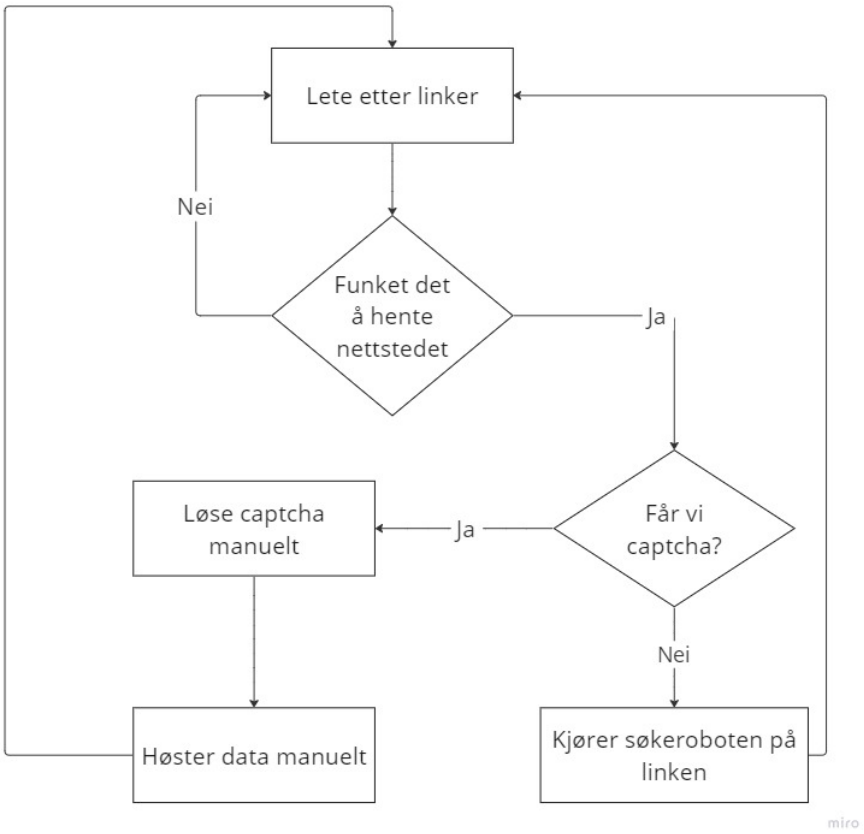
Når funksjonaliteten for kryping og høsting av tekst og bilder var implementert og testet ferdig på det åpne nettet, gikk veien videre til å teste søkeroboten på Tor. Her ble søkeroboten testet mot New York Times sin .onion side, for å teste hvordan søkeroboten oppførte seg gjennom Tor. Senere i utviklingsfasen ble søkeroboten testet mot nettstedene til løspengevirus-gruppene RagnarLocker og LockBit. Da det var data fra slike nettsider som oppdragsgiver var interessert i, se avsnitt 3.8. På slike nettsider kunne søkerobotens nedlastningsfunksjonalitet også testes, da disse nettsidene inneholdt knapper for å laste ned filer. Nettsidene til både RagnarLocker og LockBit har begge implementert sikkerhetsmekanismer mot nylige tjenestenektangrep i form av DDos-Protection og CAPTCHA. Disse sik-

kerhetsmekanismene bidro til at utviklingen av søkeroboten ble mer innviklet og den ble spesialtilpasset sider med DDoS-Protection, som vist i figur 3.6.

Figur 3.6: Utklipp av LockBit sin DDoS-Protection

3.8 Datahøstingen

Datahøstingen ble utført gjennom en blanding av automatisk og manuell høsting. Høstingen skjedde på et nettverk med Gigabit-hastighet, og gjennom en VPN-tjeneste som er kjent for å opprettholde høye hastigheter. Den automatiske høstingen ble utført av søkeroboten. Samtidig som søkeroboten kjørte, utførtes en manuell form for høsting. Prosessen for den manuelle høstingen vises i figur 3.7.



Figur 3.7: Prosess for manuell høsting

3.8.1 Manuell datahøsting

Som vist i figur 3.7, startet prosessen for datahøstingen med å finne URLer. Denne letingen etter URLer ble i hovedsak gjort på det åpne nettet, med noen unntak hvor det mørke nettet ga resultater. Nettsider som ble brukt for å lete etter URLer på det åpne nettet er Reddit og Darkfeed[2](som gruppen ble tipset om av NC3). Letingen på det mørke nettet ble gjort i form av å besøke forum, og videre ble det forsøkt å finne nye URLer gjennom disse forumene. I neste steg av høstingsprosessen ble den aktuelle URLen testet for å se om nettsiden var tilgjengelig. Hvis nettsiden var tilgjengelig, gikk prosessen videre og sjekket om nettsiden hadde en CAPTCHA. Hvis nettsiden ikke hadde en CAPTCHA kjørte søkeroboten på URLen. Hvis nettsiden hadde en CAPTCHA måtte denne løses manuelt, og deretter ble data manuelt høstet og lagt i databasen beskrevet i avsnitt 3.6.4.

[2]`https://darkfeed.io/ransomgroups/`

Kapittel 4

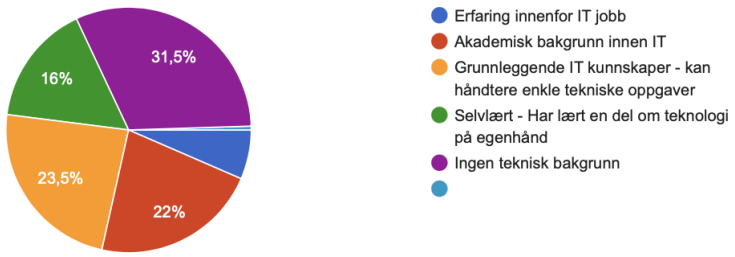
Resultater

4.1 Introduksjon

Resultatet er del inn i to deler, spørreundersøkelse og datahøsting. Resultatet og analysen av litteraturstudien er presentert i kapittel 2. Som nevnt i problemformuleringen avsnitt 1.6 ble det bestemt å bruke ulike metoder for å nå målene i denne oppgaven. Derfor vil resultatene av spørreundersøkelsen utført være en del, mens resultatene av datahøstingen gjennomført ved bruk av egen søkerobot være den andre delen av dette kapitlet.

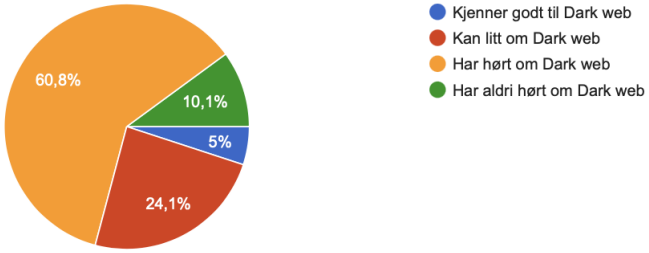
4.2 Spørreundersøkelse

Det har ikke blitt utført deskriptiv statistisk analyse av resultatene til spørreundersøkelsen. Det ble avgitt 200 svar på spørreundersøkelsen, hvorav omtrent 44 % uttrykte en form for interesse for teknologi, som oppstod som følge av studier, arbeid eller fritidsaktiviteter. Som vist i figur 4.1, har 31 % ingen teknisk bakgrunn og 23 % kan håndtere enkle tekniske oppgaver. Resultatene samsvarte ikke med forventeningene til gruppen. Det ble antatt at flere personer i gruppens nettverk hadde grunnleggende ferdigheter. Denne antagelsen var basert på gruppens erfaringer og kunnskap om sitt nettverk. Resultatet kan være påvirket av formuleringen på svaralternativene.



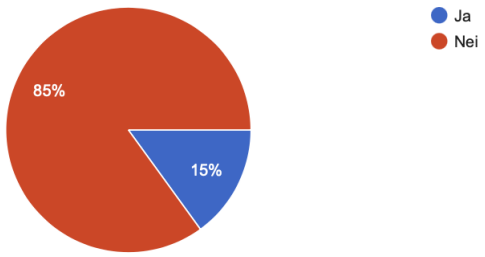
Figur 4.1: Hva er din tekniske bakgrunn?

Spørsmål nummer 2 avdekker deltakernes kjennskap til det mørke nettet. Som illustrert i figur 4.2, har godt over halvparten hørt om denne delen av internett. Det er kun 10 % som aldri har hørt om dette. På den andre siden av spekteret er det omtrent 5 % som er godt kjent med fenomenet. Det må tas høyde for at denne statistikken ikke representerer hele den norske befolkningen. Dog har gruppen kommet til enighet om å ta i bruk denne dataen på grunn av tilgjengeligheten, og at det gir en viss antydning til hva slags tanker en stor gruppe mennesker besitter om det mørke nettet.



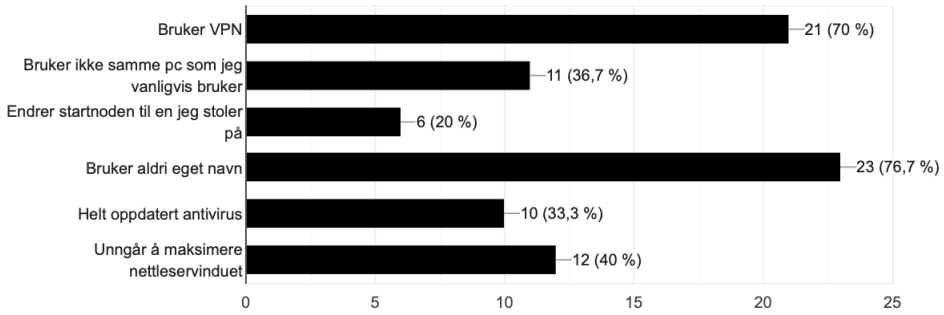
Figur 4.2: Hva er din kjennskap til dark web?

Figuren 4.3 nedenfor viser oppfølgningsspørsmålet som ble stilt. En betydelig andel av deltakerne har aldri benyttet seg av det mørke nettet, til tross for at omtrent 44 % av disse deltakerne besitter en over gjennomsnittlig teknologiforståelse. Fra spørsmål 2, se figur 4.2, har det blitt avdekket at de aller fleste kjenner til det mørke nettet, likevel er det de færreste som har turt å utforske dette på egenhånd. I dette tilfelle gjelder det kun 15 % av deltakerne.



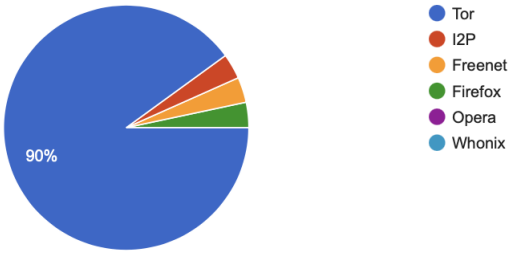
Figur 4.3: Har du noen gang brukt dark web før?

Det fjerde spørsmålet ble kun rettet mot deltakerne som hadde svart ja på spørsmål 3. Som demonstrert i figur 4.3, utgjorde denne gruppen kun 15 % av deltakerne. Spørsmålet dreide seg om forhåndsregler som kan tas for å holde seg trygg på det mørke nettet. Det var mulig for deltakerne å velge flere av svaralternativene som ble presentert. Ifølge figur 4.4 er det å bruke VPN og unngå å oppgi personopplysninger de to tiltakene som scorer høyest blant denne gruppen av deltakere. Bruk av en annen datamaskin, oppdatert antivirusprogramvare, samt å være oppmerksom på størrelsen på nettleservinduet, er alle proaktive tiltak som følger tett på hverandre når det gjelder popularitet blant deltakerne.



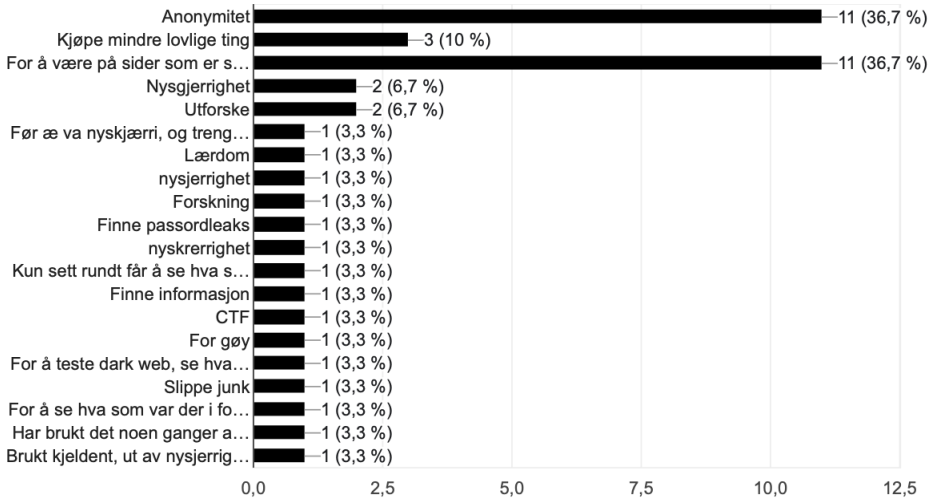
Figur 4.4: Hvilke forhåndregler tar du når du bruker dark web?

En analyse av dataene fra undersøkelsen viser at blant deltakerne som svarte på spørsmål om bruken av nettlesere for tilgang til det mørke nettet, er Tor den mest foretrukne nettleseren, som vist i figur 4.5. Dette resultatet er i tråd med den generelle populariteten til Tor som en nettleser for anonym surfing på nettet [69].



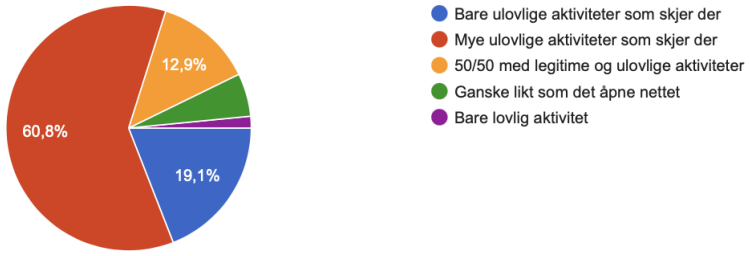
Figur 4.5: Hvilken nettleser bruker du på dark web?

Det er interessant å vite årsaken for bruk av det mørke nettet, derav spørsmålet som vises i figur 4.6. Flertallet bruker det mørke nettet på grunn av anonymitet og for å besøke nettsider som sperret eller utilgjengelig på det åpne nettet. Disse resultatene viser til at det finnes legitim bruk av Tor, og det ikke er kun ondsinnede aktører som bedriver kriminalitet.



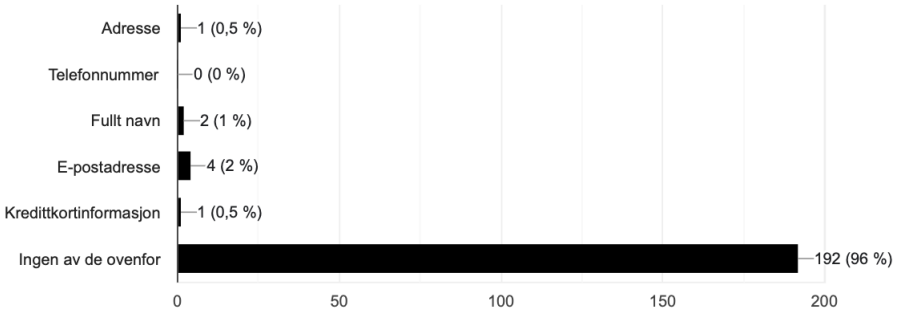
Figur 4.6: Hvorfor bruker du dark web?

I figur 4.7 presenteres resultatene av undersøkelsen som viser hvilke inntrykk deltakerne har om teknologien. Resultatene kan skyldes den negative assosiasjonen som omgir det mørke nettet. Fordommer mot denne delen av internett kan tilskrives medienes fremstilling av det, der det ofte knyttes opp mot ondsinnede aktører og skaper frykt blant publikum. Videre er det verdt å bemerke at mange ikke er kjent med Tors selvpålagte oppdrag, slik det er beskrevet i avsnitt 2.3. Omtrent 60 % av deltakerne mener at det er *mye* ulovlig aktivitet på det mørke nettet, og 19 % antar at det *bare* er ulovlig aktivitet. Dog er det en liten andel som antar at lovligheten til aktiviteten på det mørke nettet er på lik linje med det åpne nettet.



Figur 4.7: Hva er din oppfatning av lovligheten til aktivitetene som foregår på dark web?

Gruppen ønsket også å vite hva slags informasjon deltakerne som benyttet seg av det mørke nettet, var villig til å dele. Dette var av interesse for å avdekke årvåkentheten blant deltakerne. Det er ikke anbefalt å dele personopplysninger på denne delen av internett ettersom at dette tar bort hensikten med anonymiteten Tor sørger for. Basert på resultatene som vist i figur 4.8, viser dette at flertallet er kritisk til hvilke opplysninger de velger å dele på det mørke nettet.



Figur 4.8: Hvilken informasjon ville du vært villig til å dele om deg selv på dark web?

4.3 Datahøsting

En søkerobot fungerer som forklart i avsnitt 2.6, ved at den finner alle URLer fra hver enkelt nettside og går gjennom disse frem til den ikke finner flere URLer. Gjennom prosessen med datahøstingen ble gruppens mistanker styrket, i det faktum at det mørke nettet ikke henger sammen på samme måte som det åpne nettet. Søkerobotene ble kjørt fra 12. april til 10. mai. I løpet av denne tidsperioden måtte søkerobotene manuelt startes opp med nye start-URLer flere ganger.

Harddisken har blitt fylt med filer, tilsvarende 226 GB, noe som tilsvarer hele databasestrukturen. Av de totale 226 GB med data som er blitt høstet, tilsvarer dette 120 476 antall filer. Dette antallet er fordelt på 11 408 mapper. Det vil være de nedlastede filene som står for mesteparten av plassen brukt på harddisken.

4.3.1 Resultater av automatisk og manuell høsting

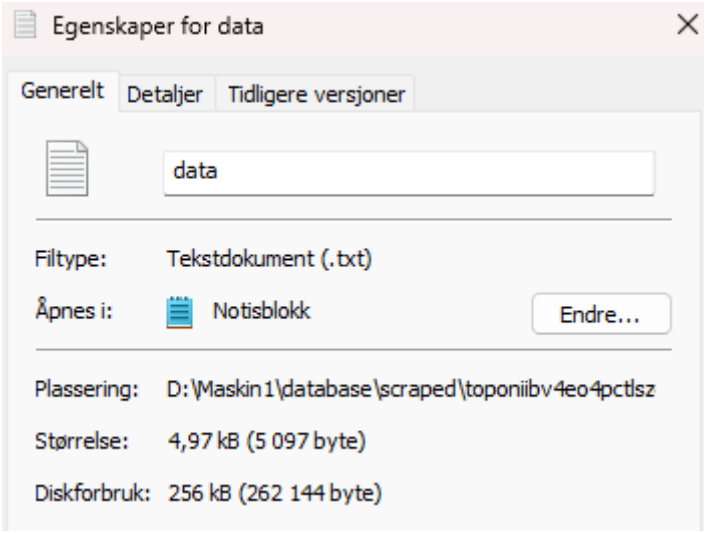
Som forklart i avsnitt 1.10.2, ble fire maskiner utdelt av Digdir og skulle være de eneste inngangspunktene til Tor-nettverket. To av disse maskinene, kalt maskin 1 og 2, kjørte søkerobot gjennom hele datahøstingen. Maskin 3 ble brukt til manuell høsting, maskin 4 ble brukt til en kombinasjon av både manuell høsting og kjøring av søkeroboten. Den totale mengden med høstet data tilsvarer ca. 199 GB med data, årsaken til at dette tallet ikke er 226 GB er forklart under tabellen. De høstede dataene er vist i avsnitt 4.3.1, der hver rad representerer hver av de fire PCene.

PC	Høstingsmetode	Total størrelse	Antall filer	Antall mapper
1	Automatisk	12,5 GB	807	1172
2	Automatisk	59,2 MB	42 746	2683
3	Manuell	170 GB	46 601	4275
4	Kombinasjon	15,9 GB	30 321	3265

Tabell 4.1: Tabell over totalt høstede data

Forskjell på minne brukt og faktisk størrelse på filer

Harddisken har blitt fylt med 226 GB med filer, men har egentlig bare 199 GB med innhold. Dette vil si at grunnet intern fragmentering på harddisken vil filene ta opp mer plass enn det filstørrelsen tilsier. Intern fragmentering kommer av at minnet er delt opp i satte blokkstørrelser [70]. Figur 4.9 viser et eksempel av en tekstfil som inneholder all tekst fra en nettside, her inneholder filen 4,97 KB med tekst, men tar opp 256 KB med minne.



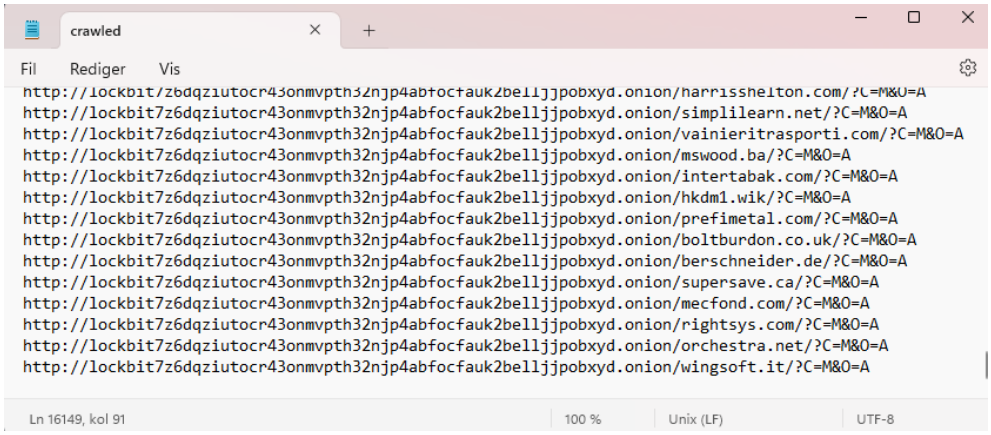
Figur 4.9: Intern fragmentering av tekstfil

Automatisk høsting

Maskin 1 er en av maskinene som har kjørt søkeroboten gjennom hele datahøstingen. Den har høstet 807 filer, som er fordelt på 1172 mapper. Årsaken til at det er flere mapper enn det er filer her, er at det er mange av nettsidene som ikke har bilder. Det betyr at søkeroboten har laget en "images" mappe, men uten innhold. Maskinen har også fått startet en nedlasting av én stor fil, det er denne som utgjør mesteparten av de 12,5 GB'ene med data. Utenom denne filen har maskin 1 høstet 26,4 MB med data.

Maskin 2 er den andre maskinen som har kjørt søkeroboten gjennom hele datahøstingen. Denne maskinen har høstet 42 745 filer, fordelt på 2683 mapper. Maskinen har ikke startet noen nedlastinger, noe som gjør at den totale størrelsen av høstede data ikke er større enn 59,2 MB. Dataene som er høstet består av tekstfiler og bilder.

Som vist i vedlegg E, vil søkeroboten legge til alle URLer i en liste, disse blir skrevet til en fil ved navn "crawled". Hver linje i denne filen tilsvarer en enkelt URL. Nedre venstre hjørne av figur 4.10 viser at maskin 2 har funnet 16 149 ulike URLer som den vil forsøke å krype gjennom, og URLene avbildet er noen eksempler på disse.



Figur 4.10: URLer funnet av maskin 2

Manuell høsting

Maskin 3 er den eneste maskinen som ikke har kjørt søkeroboten noe som helst gjennom datahøstingen. Den har kun blitt brukt til å manuelt manøvrere seg gjennom det mørke nettet og finne data. Resultatene av den manuelle høstingen har vært primært store filer og data knyttet til løsepengevirusgrupper. Disse dataene har en størrelse på 170 GB, og består av 46 601 filer, fordelt på 4275 mapper.

Kombinasjon av manuell og automatisk høsting

Maskin 4 har blitt brukt til en kombinasjon av å kjøre søkeroboten og til å manuelt manøvrere seg på det mørke nettet. På denne maskinen er det hentet data som tekst og bilder gjennom søkeroboten, mens større filer er blitt hentet ned manuelt. Gjennom maskin 4 er det blitt høstet 15,9 GB med data, tilsvarende 30 321 filer, fordelt på 3265 mapper.

4.3.2 Eksempler av funn

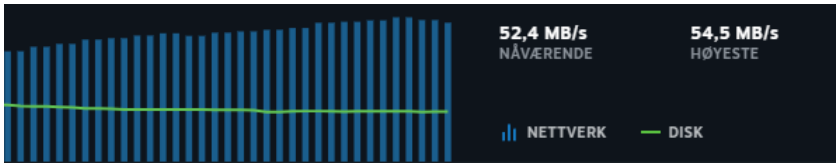
Siden den totale datahøstingen har resultert i veldig mange filer, som nevnt i avsnitt 4.3, er det ikke mulig å vise alt av data som er høstet. I dette delkapitlet vil det bli vist flere eksempler på høstet data. Eksemplene brukt er et tilfelle av én nedlasting av en stor fil på Tor-nettleseren. Videre vil det bli presentert et utvalg av URLer som søkeroboten har høstet data fra, samt mappene som er blitt laget til de utvalgte URLene. Det vil også bli vist hvordan mappestrukturen til en utvalgt URL ser ut, samt de høstede dataene til den URLen.

Nedlastingshastighet på Tor

Under er det vist eksempel av hvordan nedlastingshastigheten er på Tor-nettverket, se figur 4.11. Nedlastingen er av en komprimert mappe, i form av en .zip mappe. Dersom en komprimert mappe har en total størrelse på 3 GB, vil det bety at innholdet i mappen er større enn det som synes i komprimert format. Hastigheten på nedlastingen er i dette eksempelet 220 KB/s.

Figur 4.11: Eksempel på nedlastingshastighet på Tor

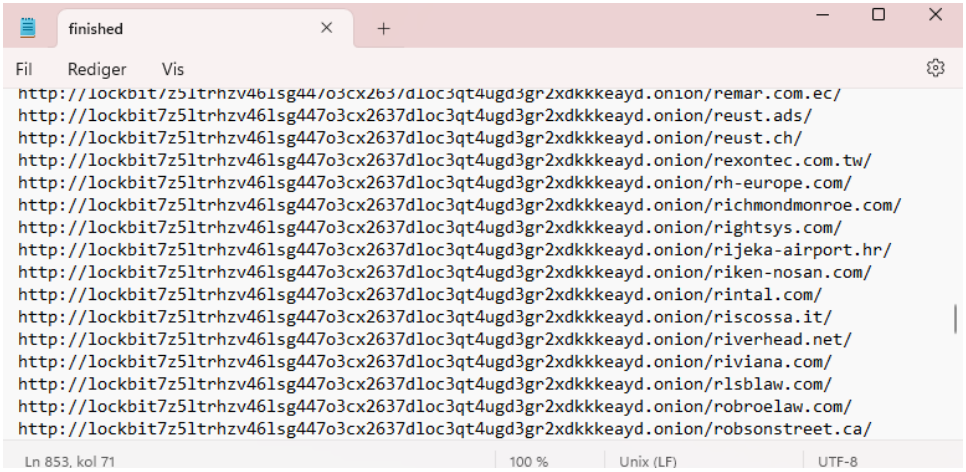
Som nevnt i avsnitt 3.8 blir nedlastingen gjort over et nettverk med Gigabit-linje. Over dette nettverket er det tidligere målt nedlastingshastigheter opp til 54,5 MB/s, som vist i figur 4.12. Dette tilsier at i eksemplet vist i figur 4.11 blir kun 0,4 % av nettverkshastigheten utnyttet.



Figur 4.12: Eksempel på nedlastingshastighet gjennom det åpne nettet

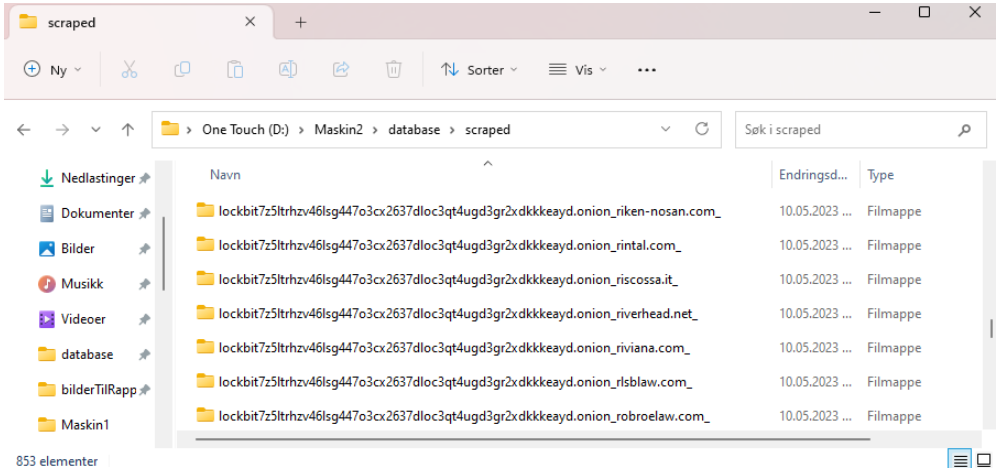
Gjennomgåtte URLer og mapper

I figur 4.13 vises gjennomgåtte URLer fra maskin 2. Kryping og skraping har blitt utført på URLene. Dette er kriteriene for at en URL er gjennomgått. Nedre venstre hjørne av figuren viser at det i løpet av høstingen er det gjennomgått 853 URLer av maskin 2. Hver av disse har blitt høstet inn i egne mapper, der de har undermapper til de forskjellige typer data som bilder, tekst og HTML.



Figur 4.13: Gjennomgåtte URLer

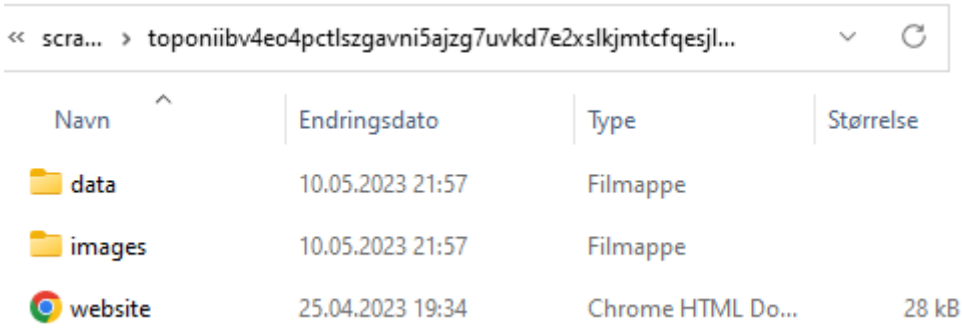
Figur 4.14 viser mappene til et utvalg av de gjennomgåtte URLene som ble presentert i figur 4.13.



Figur 4.14: Mapper til gjennomgåtte URLer

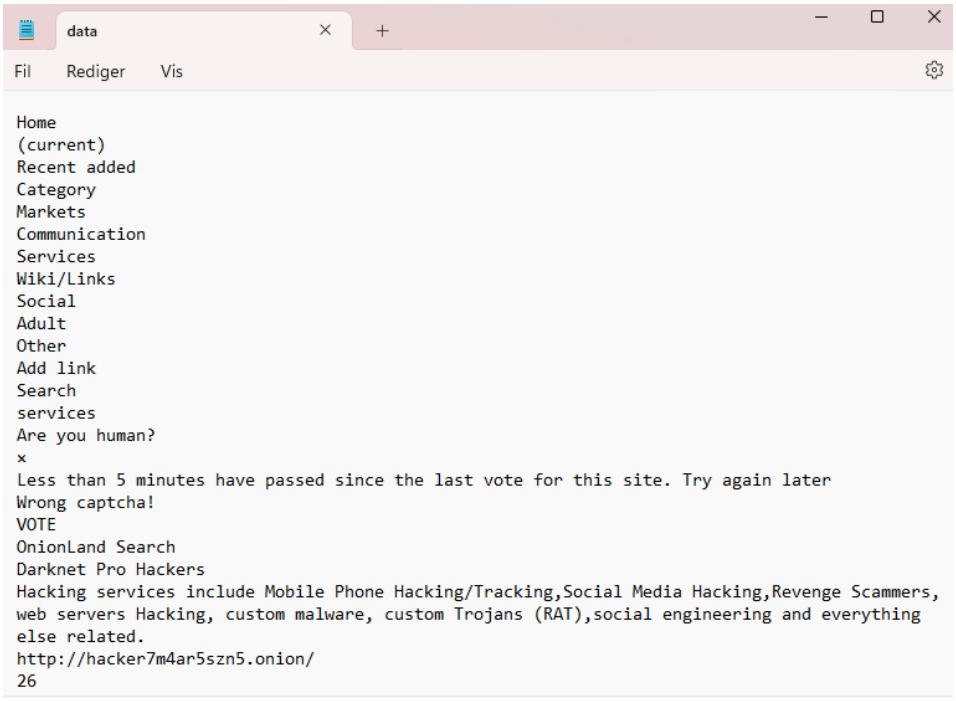
Mappestruktur og høstede data

Figur 4.15 viser mappene til de høstede datene fra en nettside på det mørke nettet. Mappen *data* inneholder en tekstfil, som er vist i figur 4.16. Mappen *images* inneholder bildene funnet på nettsiden, og er vist i figur 4.17. Årsaken til at det ikke er en mappe med navn *downloads*, var at det ikke var noen statiske filer som kunne lastes ned på nettsiden. *Websites* er kopien av nettsiden, som er vist i figur 4.18.



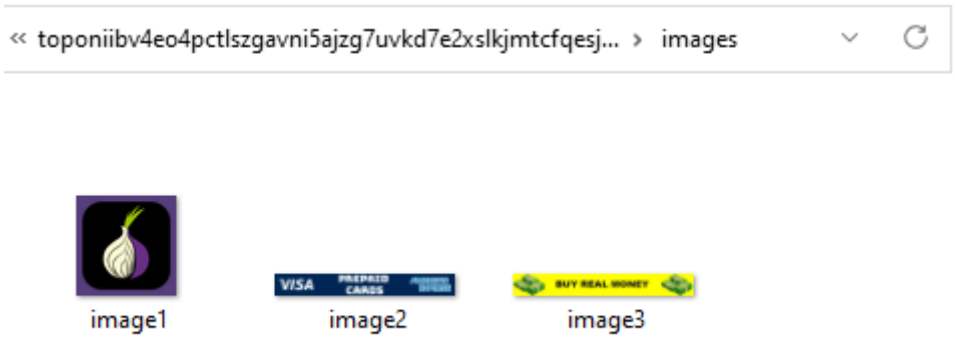
Figur 4.15: Mappestruktur fra eksempelnettside

Figur 4.16 viser hvordan den høstede teksten ser ut. Hvert tekstelement fra nettsiden vil bli skrevet på en egen linje.



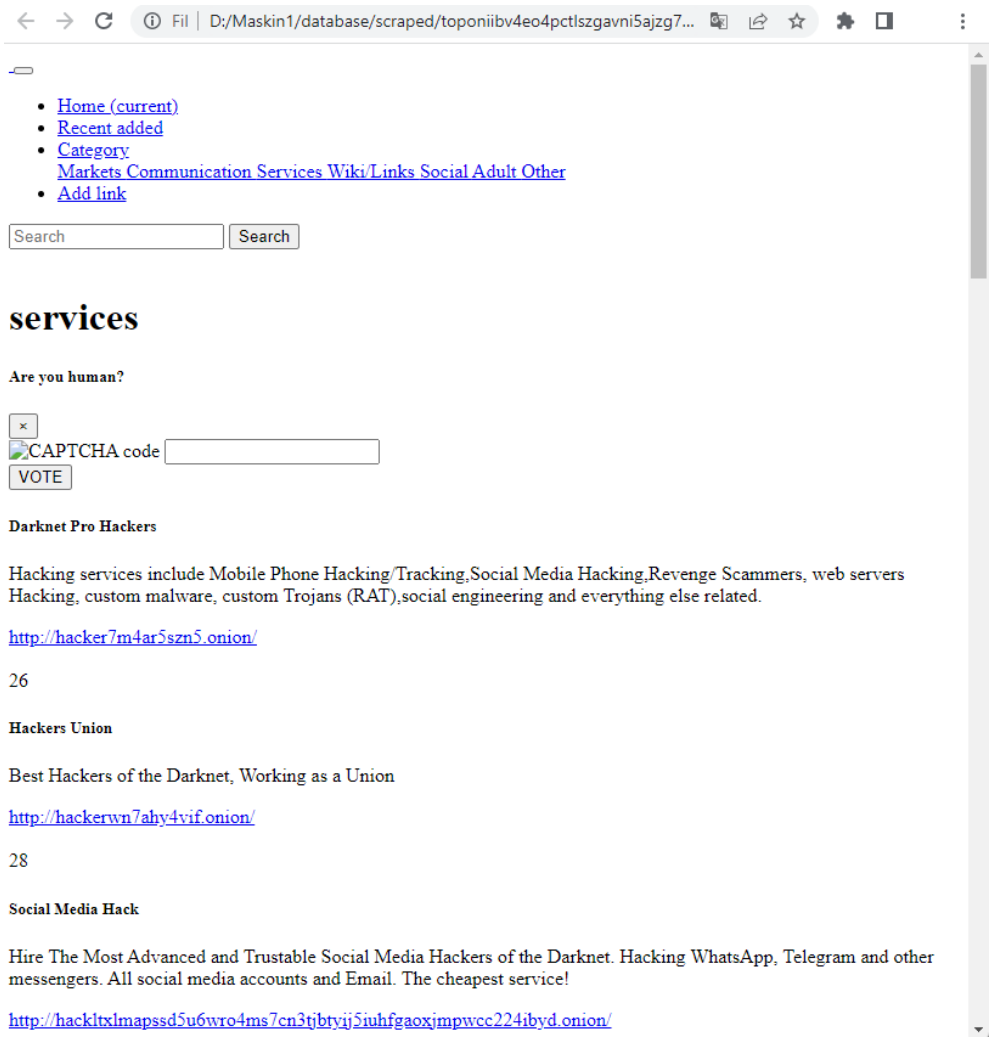
Figur 4.16: Teksten som er høstet fra nettsiden

Figur 4.17 viser bildene som er høstet fra nettsiden vist i figur 4.19. En sammenligning av den reelle nettsiden og kopien, vist i figur 4.18, viser at det ikke bare er CSS som mangler i kopien av nettsiden. Derfor høstes alle bildene med i en egen mappe i samme rekkefølge som de vises på nettsiden.



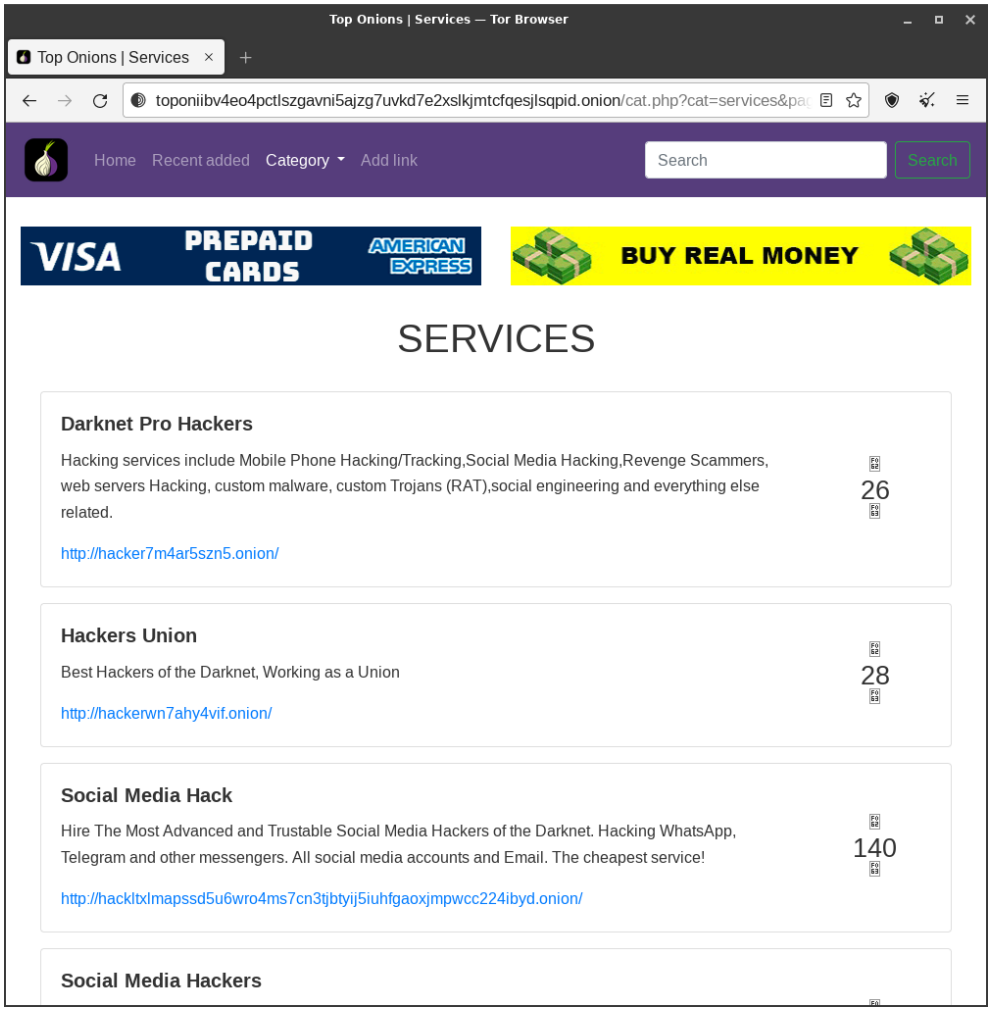
Figur 4.17: Bilder hentet fra eksempelnettside

Figur 4.18 viser et eksempel på hvordan en kopi av en nettside som er blitt høstet av søkeroboten ser ut. Kopien av nettsiden viser at det er programmert inn en CAPTCHA, men denne har søkeroboten klart å unngå.



Figur 4.18: Kopi av nettside fra det mørke nettet

Nettsiden det er høstet en kopi av, er vist i figur 4.19. Årsaken til at kopien av nettsiden ikke ser ut som den originale, er fordi kopien kun inneholder HTML, og ikke har med CSS. I all hovedsak er dette generelt design og bilder på nettsiden. Dette kan en se ved å sammenligne figur 4.18 og figur 4.19. Kopien av nettsiden i figuren over, viser til at det er en CAPTCHA tilstede. Dersom søkeroboten ikke hadde programmert inn ventetider mellom krypingen, ville denne CAPTCHAen mest sannsynlig kommet opp etter hvert, og data høstet fra flere nettsider ville blitt ubrukelige.



Figur 4.19: Nettside fra det mørke nettet

Kapittel 5

Diskusjon

5.1 Introduksjon

Dette kapitlet drøfter og sammenligner utfallet av metodene brukt for å nå målene beskrevet i avsnitt 1.6. Det vil også diskuteres avgjørelser som er tatt underveis, erfaringer gruppen har fått i løpet av prosessen, hvilke begrensninger og utfordringer som har påvirket arbeidet, og videre arbeid.

5.2 Begrensninger

Datahøsting

Data høstet er begrenset til data hentet i perioden 12.april til 10.mai. Det er kun data fra denne perioden som er lastet ned og tatt i betrakning i denne oppgaven. Det vil si at sider som har vært utilgjengelig på Tor i den tiden ikke har vært mulig å høste data fra.

V2 og V3 onion-lenker

Som nevnt i avsnitt 2.5.2 er det kun V3 onion-lenker som fungerer på Tor nå. Tidligere var det V2 onion-lenker, men selv om disse ikke fungerer så eksisterer de fortsatt på Tor. Eksempel på disse to typene lenker kan sees på figur 4.19, der V2 er de to korte og V3 er den lengste. Det at V2 onion-lenker fortsatt ligger på Tor gjør at søkeroboten vil bruke tid på å prøve å nå disse lenkene, uten at det vil gi noen resultater. Dette viste seg i ettertid å kunne løses på en grei måte, noe gruppen ikke fant ut av før det var for sent. Gruppen måtte generelt forholde seg til den versjonen av Tor som var tilgjengelig i månedene prosjektet varte.

Statistikk

Statistikk om bruken og oppfatningen av det mørke nettet ble funnet på Statista.no, men denne informasjonen var begrenset av en betalingsmur. Tilleggsdata

er viktig for å kunne trekke pålitelige konklusjoner og gi tilstrekkelig informasjon. Gruppen løste dette på en annen måte, ved å ta i bruk undersøkelsen utført av CIGI. Den supplerte gruppens egen spørreundersøkelse, selv om det i tillegg hadde vært nyttig å ha statistikken fra Statista også.

5.3 Resultater

Det teoretiske fundamentet for oppgaven består av en litteraturstudie om Tors bakgrunn, historie, fordelaktig bruk og den tekniske oppbyggingen. I tilegg til informasjon om egenutviklet søkerobot.

Som beskrevet i avsnitt 2.1.1 finnes det ulike nivåer av internett. Tor er en del av det mørke nettet og det er viktig å få frem at det ikke er tilgjengelig på den delen av internett som folk flest bruker. Det mørke nettet er en del av det dype nettet, men for å nå dit trengs en spesiell programvare og det er viktig å påpeke at en ikke havner på Tor med et uhell. Det er noe som blir gjort med viten og vilje, men å laste ned programvaren Tor og koble seg på Tor-nettleseren i seg selv, er ikke ulovlig. Tor ble som nevnt i avsnitt 2.2 hovedsaklig utviklet for at amerikansk etterretning skulle ha tilgang til å kommunisere anonymt på nett, men ble fort overtatt av brukere som drev ulovlig virksomhet på grunn av mulighetene for å holde seg anonym. Dette kan en nok si var starten på det synet allmennheten har på det mørke nettet. Betegnelsen "dark web" og manglende tilgjengelighet via det åpne nettet, kombinert med anonymiteten som overholdes, er sannsynligvis med på å bidra til den negative assosiasjonen folk forbinder med det mørke nettet [14].

Utforskingen av fordelene ved Tor, presentert i avsnitt 2.3, fremhever områder som samsvarer med Tor Prosjektets selvpålagte oppdrag. Det er av interesse å utforske hvilke brukere og bruksområder som samsvarer med målet til Tor Prosjektet, med sikte på å bedre forstå bruken av nettverket. Flere områder var ofte gjentakende, og gruppen konsentrerte seg om enkelte av disse, hvor det ble identifisert hendelser som understøttet argumentet for fordelene med Tor.

Eksempelvis blir menneskerettighetsaktivister anbefalt å benytte seg av det mørke nettet for å beskytte seg selv når de kommuniserer ut informasjon. Dette fordi flere land ikke følger menneskerettighetene og det å stå opp for det og være i mot regimet både er farlig og forbudt enkelte steder. På en side kan en jo si at det de gjør er ulovlig, fordi landet de bor i ikke tillater slike ytringer, men på den andre siden er det snakk om at disse landene bryter menneskerettighetene. Det kan derfor være vanskelig å vurdere om en kan kalle det legitim bruk av Tor, da det avhenger av hvilket perspektiv en ser det fra. Det vil likevel anses av gruppen som fordelaktig bruk av Tor, og mye av grunnen til at Tor eksisterer, ifølge dem selv.

Et annet nevnt eksempel er den arabiske våren, beskrevet under avsnitt 2.3.1. Dette er sannsynligvis noe flere er kjent med, likevel er det kanskje ikke kjent

at Tor spilte en rolle under den perioden. I flere av landene som ble rammet av opprøret, valgte regimene som nevnt å stenge internett for å hindre innbyggerne i å avtale demonstrasjoner og dele og lese nyheter. Når det gjaldt situasjonen i Egypt, kan det diskuteres om nedstengningen av internett faktisk førte til en økning i antall deltakere i demonstrasjonene, da det kan ha tvunget innbyggerne ut på gatene for å skaffe informasjon. Enkeltpersoner som ble angrepet digitalt på grunn av deres motstand, fikk ved bruk av Tor beskyttelse av sin identitet. Dette kan ha ført til at flere turte å ytre sin mening, da de ikke lengre trengte å være redd for personlige konsekvenser for seg eller sin familie. På denne måten var Tor relevant i en historiske hendelse, noe som er med på å illustrere fordelaktig bruk av Tor.

En fordel med Tor, som ikke nødvendigvis er bruksområder av Tor, er at formålet ikke er å tjene penger. Dette medfører at myndigheter ikke har en påvirkning på tjenestene deres, forklart i avsnitt 2.3.2. Sett fra et annet perspektiv kan en drøfte ulempene ved at det ikke er en myndighet som håndhever lovene på Tor-nettverket. Retningslinjer burde kanskje bli mer aktivt håndtert på det mørke nettet. Det har i flere tilfeller blitt tatt ned ulovlige markeder og pedofilringer på det mørke nettet, noe som viser at det faktisk er mulig å få til. Dersom en hadde hatt en type myndighet som sørget for å håndheve lover eller retningslinjer satt, vil en kanskje få bort den ondsinnede aktiviteten som foregår, men samtidig beholde fordelene. Det vedvarende spørsmålet som gjentar seg i denne diskusjonen er om fordelene veier opp for ulempene.

Sammenligning av funn spørreundersøkelser

Spørreundersøkelsen som ble sendt ut til gruppens sosiale netteverk var noe forhastet. I det innledende spørsmålet ble deltakerne presentert for fem forskjellige svaralternativer som vist i figur 4.1, og det er grunn til å anta at "grunnleggende IT kunnskaper" kan ha blitt tolket feil av flere. Kategorien "grunnleggende IT kunnskaper" skulle inkludere alle som klarte å utføre enkle tekniske oppgaver, som å gjenopprette en datamaskin. Derfor er det sannsynlig at flere av deltakerne som oppga at de ikke har teknisk bakgrunn likevel besitter grunnleggende IT-kunnskaper. Som tidligere nevnt, har flere av spørsmålene i spørreundersøkelsen blitt presentert med svaralternativer, og dette kan også vært med på å påvirke resultatene av undersøkelsen. Valget landet på flervalg for enkelthetensskyld, etter egen erfaringer er terskelen for å svare på spørreundersøkelser med svaralternativer mye lavere enn å svare på spørreundersøkelser med fritekst. Formålet med spørreundersøkelsen var å få mange nok besvarelser slik at det var mulig å trekke konklusjoner basert på data som har blitt innhentet av gruppen selv. Svaralternativene som blir illustrert legger føringer på hva deltakerne kan svare, og det var kun et av spørsmålene som gjorde det mulig for deltakerne og legge til egne svar. Denne utformingen kan derfor ha påvirket resultatene til spørreundersøkel-

sen negativt.

I avsnitt 4.2 ble det belyst at gruppen har valgt å bruke resultatene fra en spørreundersøkelse utført at CIGI som et supplement, da dette er en undersøkelse utført globalt på en større folkegruppe. Til tross for en større andel svarprosent er resultatene fra begge undersøkelsene relativt samstemte. Statistikken varierer fra land til land men det har blitt tatt utgangspunkt i gjennomsnittet uavhengig av geografisk lokasjon. Det har ikke blitt foretatt en vurdering av fordelingen i de ulike landene individuelt ettersom det ikke er relevant for problemstillingen i denne oppgaven. Tatt i betrakning at undersøkelsen til CIGI ble utført for en tid tilbake, kan det ikke fastslås med sikkerhet om resultatet ville vært det samme dersom undersøkelsen ble utført i dag. Likevel brukes resultatene fra denne spørreundersøkelsen for å understøtte gruppens resultater, ettersom dette er en av de få undersøkelsene utført av en troverdig aktør, hvor resultatene har blitt publisert på det åpne nettet.

Resultatene fra spørreundersøkelsen til CIGI avdekker at det kun er 20 % av verdensbefolkningen som har noe kjennskap til det mørke nettet, og 4 % som er godt kjent med denne delen av internett [31]. Disse resultatene stemmer overens med resultatene fra gruppens spørreundesøkelsen som presentert i figur 4.2. Basert på dette er det mulig å trekke konklusjoner om global kunnskap knyttet til teknologien.

Ved å sammenligne resultatene fra spørsmålet om hvorvidt deltakerne hadde benyttet seg av det mørke nettet fra CIGI med figur 4.3, kan det observeres at prosentandelen omtrent er den samme, til tross for at antall deltakere er forskjellig. Disse resultatene viser at selv om det mørke nettet er kjent for mange, er det likevel svært få som har valgt å utforske denne teknologien. Dette kan være knyttet opp mot resultatene fra spørsmålet om oppfatningen deltakerne besitter om det mørke nettet. Resultatene fra undersøkelsen indikerer at en betydelig andel av deltakerne, nemlig 80 %, tror at det foregår enten *kun* ulovlig aktivitet eller *mye* ulovlig aktivitet, som illustrert i figur 4.7. Spørreundersøkelsen til CIGI inneholder et tilsvarende spørsmål, spørsmålet er knyttet til manglende bruk av det mørke nettet. Som nevnt i avsnitt 2.4.1, rapporterte 46 % at de ikke vet hvordan Tor og det mørke nettet aksesseres, og 45 % ser ikke på det mørke nettet som en nødvendighet [31]. På den andre siden hevder 27 % at det mørke nettet burde være ulovlig eller at det kun blir brukt av kriminelle [31]. Disse funnene indikerer en betydelig grad av uvitenhet, utilgjengelighet og negative assosiasjoner knyttet til Tor og det mørke nettet blant deltakerne i spørreundersøkelsen.

Blant deltakerne som benytter seg av Tor, er anonymitet den primære årsaken til at dette verktøyet blir benyttet. En annen signifikant årsak er å oppnå tilgang til innhold som er blokkert eller utilgjengelig på det åpne nettet, som demonstrert i figur 4.6. Figur 2.5 fremlegger også at en stor andel av deltakerne bruker Tor eller det mørke nettet grunnet anonymitet [31]. I tillegg til dette dokumen-

terer også resultatene til CIGI at en viss andel av deltakerne bruker det mørke nettet til å få tilgang til innhold som ville vært utilgjengelig på det åpne nettet. Det er imidlertid utfordrende å avgjøre lovligheten av slik bruk, da innholdet som aksesseres er begrenset. Videre kan det være en utfordring å vurdere deltakernes oppriktighet, til tross for at begge spørreundersøkelsene er gjennomført anonymt.

Sammenligning av funn fra høstet data

Som forklart i avsnitt 5.3 finnes det mange positive aspekter ved Tor, men via spørreundersøkelsen kommer det frem at 80 % tror at det i all hovedsak er ulovlig aktivitet som skjer på Tor. Under avsnitt 5.3 er det forklart hvordan spørreundersøkelsen utført av CIGI avdekker at store deler av befolkningen ikke ser nytten av Tor, eller synes at Tor bør være ulovlig grunnet kun kriminell aktivitet. Data som er blitt hentet ut av søkeroboten støtter opp påstandene om mye kriminell aktivitet, denne dataen er preget av gruppens mål om å finne aktivitet knyttet til hackergrupper. Søkeroboten trenger som forklart i avsnitt 3.6.1, en tekstfil med en eller flere URLer som den skal starte krypingen fra. Dermed vil gruppens innblanding i søkerobotens prosess ha stor innvirkning på hvilke typer nettsteder den besøker.

I avsnitt 4.3.1 vises det hvordan maskin 1 fikk høstet 807 filer, mens maskin 2 klarte å høste 42 746 filer i løpet av samme tidsperiode. Som nevnt under avsnitt 5.2, ligger det enda V2 lenker ute på noen nettsider på Tor-nettverket. Slike lenker ble funnet av maskin 1, som vist i figur 4.18. Dette kan forklare hvorfor det ble høstet færre filer på maskin 1, enn det ble på maskin 2. Den eneste forskjellen på skrapingen gjort av maskinene, var hvilken URL de skulle starte fra, altså tekstfilen som skulle aksesseres av søkeroboten. Maskin 1 ble i tillegg stoppet manuelt på flere tidspunkt fordi sidene den skulle til å gå inn på, var sider som gruppen ikke var interessert i. Som forklart i avsnitt 2.5.2, så er nettsidene på Tor kjent for å dukke opp, og forsvinne like raskt igjen. Dette førte til at maskinene som kjørte søkeroboten, prøvde å sende forespørsler til flere nettsider som ikke lengre eksisterte. Dermed ble mye av tiden søkerobotene kjørte, brukt til dette.

Hvis en ser på forskjellene på resultatene fra den manuelle høstingen kontra den automatiske høstingen, som presentert i avsnitt 4.3.1. Den manuelle høstingen resulterte i ca. 170 GB med data, kontra den automatiske høstingen som samlet resulterte i ca. 12,5 GB med data. Den store størrelsesforskjellen har nok grunn i at den manuelle høstingen gir mye større resultater enn den automatiske høstingen. Den automatiske høstingen har også hatt vanskeligheter rundt nedlasting av filer, som er beskrevet i avsnitt 5.4.4. Resultatene av kombinasjonen av høstingsmetodene, presentert i avsnitt 4.3.1, har gitt 3 GB mer data enn kun automatisk høsting. Det er den rene manuelle høstingen som dominerer når det kommer til høsting av data. Grunnen til dette er nedlastinger av store filer, og at det ikke er nødvendig å lete gjennom alle lenkene som er tilgjengelige på nettsider, som

søkeroboten er programert til å gjøre.

Veier det positive opp for det negative?

I figur 2.6 fremlegges det at 66 % av verdensbefolkningen mener at det mørke nettet bør legges ned [31]. Dette omdømmet er ikke helt ufortjent, for det er ingen tvil om at denne delen av internett er preget av ulovlige aktiviteter, som blant annet salg av rus, våpen, stjålne identiteter og barnepornografi. Ved hjelp av navigasjon og utforskning har gruppen selv avdekket de mørke sidene ved Tor. Konklusjonen som ble trukket etter en relativ lang prosess med etterforskning, er at trafikken på det mørke nettet ofte karakteriseres som ulovlig eller tvilsom. Denne konklusjonen samsvarte også med NC3(se vedlegg G) etterforskernes funn. Videre kan denne oppfatningen ha blitt forsterket blant verdensbefolkningen som følge av pressens synsvinkel. Ved å gjennomføre et søk på Google med "det mørke nettet" som søkeord er det mulig å se at det ofte er de verste sakene som skaper de største overskriftene. De første søkresultatene som dukker opp omfatter overgrepsforum, narkotikamarked og kriminalitet. Denne fremstilling kan bidra til å skape frykt og avsky blant allmennheten, og dermed påvirke oppfatningen om det mørke nettet som en farlig og kriminell arena. Det er imidlertid viktig å merke seg at det mørke nettet også brukes til positive formål som presentert i avsnitt 2.3, Tor og det mørke nettet har hatt positiv innvirkning på verdensbefolkningen ved å fremme ytringsfrihet og kampen mot internettsensur. Det store spørsmålet er om de positive aspektene veier opp for de negative? Dette er helt klart et vanskelig spørsmål å svare på ettersom det finnes positive og negative sider ved alt. I avsnitt 2.3 blir det nevnt tilfeller, som den arabiske våren, der Tor gjorde det mulig for opprørerne å få avtalt demonstrasjoner samt dele nyheter. Dette er noe som ikke hadde vært gjennomførbart uten det mørke nettet grunnet regimene som fjernet innbyggernes kobling til internett. Et annet, nyere eksempel vil være internettsensuren pålagt av russiske myndigheter. Ved hjelp av Tor Prosjektet som oppgraderte programvaren deres ved å implementere en ny funksjon som gjorde det mulig å komme seg forbi blokkeringen satt opp av myndighetenere, har gjort det mulig for den russiske befolkningen å få tilgang til det frie internettet. Derfor er det nødvendig å vurdere pressens presentasjon av det mørke nettet i en bredere kontekst, og ta hensyn til både positive og negative aspekter ved denne teknologien.

5.4 Utfordringer underveis

Dette underkapitlet vil beskrive utfordringer gruppen opplevde underveis, som ikke kunne forutses på forhånd og som har gjort at resultatet ikke ble som gruppen hadde planlagt.

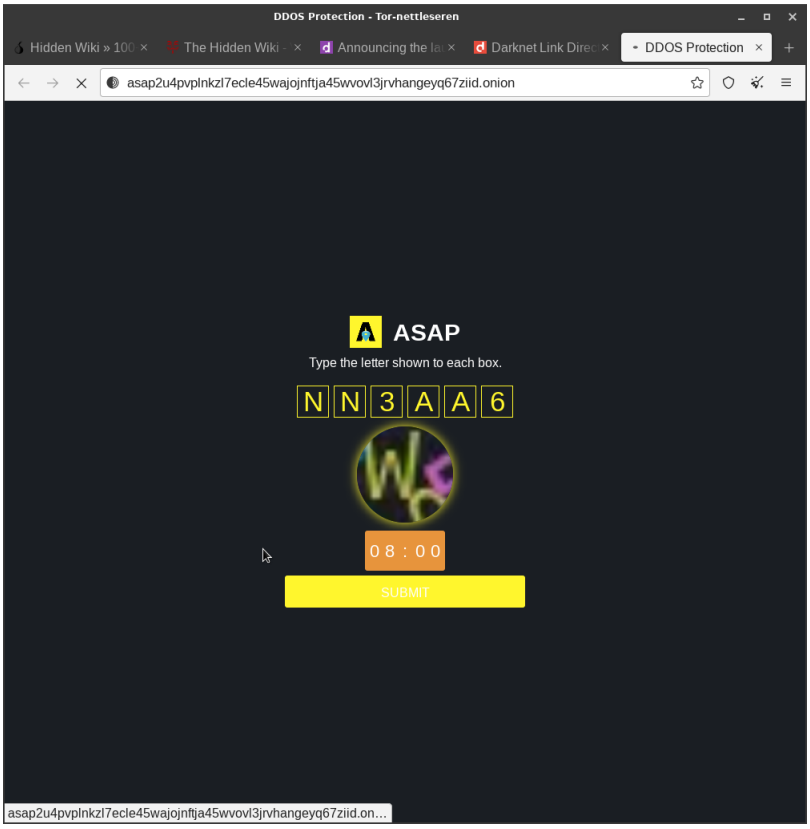
5.4.1 Prosess i arbeid med søkeroboten

Gjennom utviklingen, testingen og kjøringen av søkeroboten har det dukket opp flere utfordringer som måtte tas hensyn til. Videre førte dette til at utvikling, testing og kjøringen ble mer tidkrevende. Utfordringene som omhandler søkeroboten er uventede sikkerhetsmekanismer på nettsider på det mørke nettet, samt programmeringsproblemer og problemer knyttet til høsting av data.

5.4.2 Sikkerhetsmekanismer på det mørke nettet

Utfordringene med CAPTCHA

I testfasen av søkeroboten ble det observert at flere av nettsidene på det mørke nettet hadde implementert sikkerhetsmekanismen CAPTCHA. Dette gjør at brukeren må løse en oppgave, enten om det er i form av en tekst eller å velge bilder av for eksempel trafikklys. Figur 5.1 viser et eksempel av en tekstbasert CAPTCHA som krever at brukeren skriver inn tegnet som er vist i sirkelen. Figuren viser til sikkerhetsmekanismen til markedsplassen ASAP market.



Figur 5.1: Eksempel av en tekstbasert CAPTCHA

For at søkeroboten skal kunne løse oppgavene presentert gjennom CAPTCHA, krever dette komplisert og tidkrevende programmering. Dersom disse oppgavene skal løses automatisk av søkeroboten, kan det antas at den er nødt til å ha kunstig intelligens implementert. Gruppens løsning av denne utfordringen var å løse disse oppgavene manuelt. Problemet som oppstod ved å løse CAPTCHA manuelt, var at søkeroboten ikke ville få tilgang til nettsiden som presenterte en CAPTCHA-oppgave. Dermed vil dataene som vil bli høstet fra en CAPTCHA-side, være selve nettsiden som har CAPTCHAen. Følgene av dette vil totalt sett være at maskinene som har kjørt søkeroboten ikke har høstet data fra nettsteder som har CAPTCHA implementert.

For å kunne ha løst CAPTCHA-problemet bedre, kunne en ha plassert URLene til nettsidene med CAPTCHA i en egen mappe eller et eget tekstdokument. På denne måten vil gruppen ha oversikt over alle nettsider med CAPTCHA som søkeroboten har kommet over, men som ikke har blitt høstet data fra.

Utfordringene med innlogging på nettsider

En annen observasjon som ble gjort i testfasen av søkeroboten var at flere nettsider hadde et krav om innlogging for å få tilgang til nettstedet. Nettsteder som har blitt observert til å kreve innlogging er diverse forum-sider, markedsplasser og chatterom. Disse nettsidene har ikke blitt inkludert i den automatiske høstingsprosessen, med mesteparten av årsaken liggende på vanskeligheter i hvordan søkeroboten skal behandle tilfeller hvor den møter på innloggingssider.

En potensiell løsning av problemet med innloggingssider gruppen implementerte, var å prøve og logge inn i på nettsidene ved bruk av tilgjengelig funksjonalitet i Selenium. Dette innebar å sende med et brukernavn og passord til søkeroboten. Kodeliste 5.1 viser hvordan denne funksjonen ville vært programmert. Den sender med *StanleyTheCatMan* som brukernavn, og *EnToTre123* som passord, for så å trykke "enter". Brukernavnet var tilfeldig valgt og kunne ikke kobles mot gruppens medlemmer.

Kodeliste 5.1: Funksjon til å logge inn på innloggingssider

```
def loginSelenium(driver):
    driver.find_element(By.NAME, "username").send_keys("StanleyTheCatMan")
    driver.find_element(By.NAME, "password").send_keys("EnToTre123" + Keys.ENTER)
```

Selv om denne funksjonen fungerte gjennom testingen, ble det ikke brukt gjennom kjøringen av søkeroboten. Årsaken for dette var at det ikke var noen åpenbar måte å finne ut av om at nettsiden søkeroboten befant seg på, var en innloggingsside. Det var i tillegg flere av disse nettstedene som krevde at bruker måtte lages i forkant.

Tjenestenektangrep på Tor

I løpet av de siste 7 månedene har Tor vært utsatt for flere forskjellige typer tjenestenektangrep. Disse angrepene har medført lav innlastningshastighet, samt manglende tilgang til onion-domener [71]. Det er vanskelig å evaluere omfanget av tjenestenektangrep grunnet begrensede og motstridende kilder. Det er få nettsider som offentliggjør detaljer knyttet til angrepet, grunnet frykten for ytterligere angrep. Sett fra et teknisk perspektiv, kan et tjenestenektangrep påvirke Tor-nettverket på tre ulike måter. Den første måten vil være ved å overbelaste en node med mye trafikk, slik at dette fører til nedetid som videre hindrer trafikkflyten. Blokkering av tjenester slik at brukere ikke får tilgang er den andre måten. Den siste måten er blokkering av utgangsnoder, dette vil gjøre det utfordrende for brukere å få tilgang til innhold utenfor Tor-nettverket. Tor Prosjektet jobber med kontinuerlig forbedring av nettverkets sikkerhet ved å legge til flere noder i nettverket, og ved å forbedre teknologien som brukes for å beskytte nettverket vil det også bli mer motstandsdyktig mot slike tjenestenektangrep.

Flere nettsider på det mørke nettet har implementert et sikkerhetstiltak mot nylige tjenestenektangrep, dette sikkerhetstiltaket kalles for DDoS-Protection. Som tidligere nevnt i avsnitt 3.7, og vist i figur 3.6, er dette et sikkerhetstiltak som gjør at nettsiden krever enda mer tid for å komme seg til selve nettsiden. Eksempelet vist i figur 3.6, krever at besøkeren av nettsiden må vente i en viss tid før brukeren sendes videre til hovedsiden. Løsningen gruppen kom med for denne utfordringen var å få søkeroboten til å vente ut tiden som nettsiden krevde. På denne måten kunne søkeroboten lett komme seg rundt denne utfordringen.

5.4.3 Programmeringsutfordringer

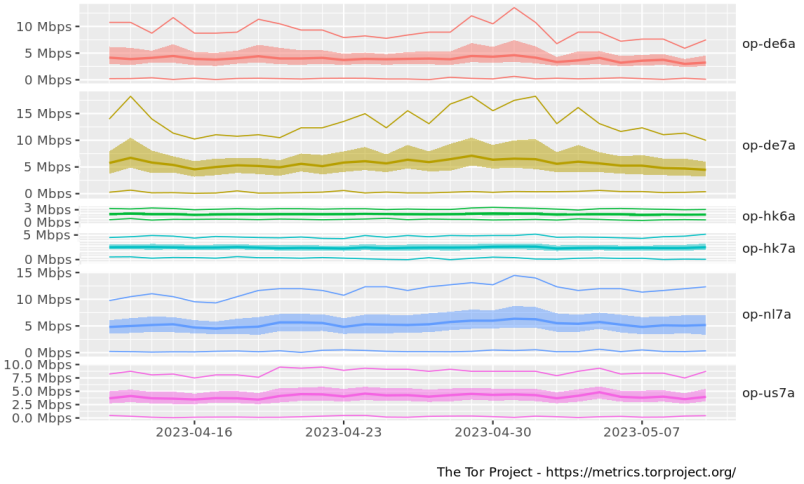
Helt fra starten av oppgaven stod det mellom to forskjellige bibliotek som søkeroboten skulle utvikles rundt. De to bibliotekene var Requests og Selenium. Selenium gir større tilgang til funksjoner som kan påvirke og interagere med nettsider, enn det Requests gjør. Requests har et større fokus på hvordan HTTP-forespørsler kan sendes via Python. Valget falt først på å bruke Requests, som har blitt forklart i avsnitt 3.6.2. Requests førte til vanskeligheter rundt bruken på Tor, noe som resulterte i uklare error-meldinger. Ved dette ble valget av hovedbibliotek endret fra Requests til Selenium, noe som førte til at utviklingsprosessen måtte startes fra bunnen. Denne endringen bidro til at utviklingen av søkeroboten tok lengre tid enn planlagt. Selenium endte opp med å ha en større nytteverdi for utviklingen, grunnet en større verktøykasse av funksjonalitet som ga søkeroboten mulighet til å oppføre seg mer som et menneske.

5.4.4 Høstingsproblemer

Å hente ned data fra det mørke nettet er en god del vanskeligere enn på det åpne nettet. Som vist i figur 4.10, ble det funnet 16 149 forskjellige URLer av maskin 2. Maskinen hadde funnet disse URLer over en periode på ni døgn med kontinuerlig kjøring. Som vist i avsnitt 3.7, klarte søkeroboten å finne over 5492 unike URLer på under én time med høsting på det åpne nettet. Ut fra dette kan det antas at det ville tatt søkeroboten cirka tre timer på det åpne nettet for å finne like mange URLer som den fant i løpet av ni døgn på det mørke nettet. Det er flere grunner til dette. Testen som ble kjørt på det åpne nettet var gjort av søkerobot versjon 1, som vist i avsnitt 3.6.2. Den første versjonen var enda ikke tilpasset for kryping på det mørke nettet i form av venting for å unngå både CAPTCHA og DDoS-protection.

Problemer med innlastingslastehastighet og gjennomstrømmingshastighet

En annen faktor som fører til at søkeroboter er mindre effektive over det mørke nettet er innlastingshastigheten av nettsider, og den generelle gjennomstrømmingshastigheten over Tor ved nedlasting av filer. Figur 5.2 viser hastigheten over Tor-nettverket ved nedlastinger av statiske filer. De høyeste og laveste målte gjennomstrømmingshastighetene vises som tynne linjer. Dette ekskluderer ekstremverdier som blir målt. Median gjennomstrømmingshastighet blir vist som litt tykkere, mørke linjer. De tykkeste linjene viser hva gjennomstrømmingshastigheten var i 50 % av tilfellene [72]. Som forklart i avsnitt 2.5.3 vil mellomnoder og sluttnoder bli tilfeldig valgt. Dette vil si at gjennom datahøstingen fikk maskinene flere ulike ruter, alle med varierende hastigheter. Som vist i figur 5.2, varierte hastighetene fra tilnærmet 0 Megabit per sekund (Mbps) og opp til rett over 18 Mbps [72]. Gjennomstrømmingshastigheten her stemmer godt med observasjoner gjort under datahøstingen, der det ble observert nedlastingshastigheter fra 19.4 KB/s og opp til 896 KB/s.



Figur 5.2: Gjennomstrømmingshastighet på Tor-nettverket ved nedlasting av statiske filer[1]

Problemer ved brudd på nettlesersesjon og brudd på nedlasting

Gjennom testfasen ble det observert at søkeroboten stoppet etter den hadde kjørt gjennom tre URLer. Det ble ikke funnet informasjon som kunne løse dette problemet ved søk på det åpne nettet. Da det ikke ble funnet en løsning på internett, ble problemet løst ved å starte en ny Selenium-driver. Med denne løsningen vil en ny Selenium-driver starte etter hver tredje URL som er besøkt, noe som betyr at søkeroboten må lukke det aktuelle nettleservinduet. Søkeroboten vil deretter starte et nytt nettleservindu, og koble seg til Tor-nettverket igjen. Denne løsningen vil føre til at søkeroboten bruker ekstra tid på å høste data.

Det ble også observert at filer som var under nedlastingsprosessen ble avbrutt. Dette har ført til at flere filer som har startet nedlasting, ikke har blitt fullførte, og dermed brukt mye unødvendige ressurser fra høstingsprosessen. Det fins måter å gjenoppta nedlastinger på Tor-nettverket, men ingen måter å gjenoppta nedlastinger som har blitt startet gjennom søkeroboten. Løsningen som ble brukt for å gjenoppta manuelle nedlastinger krevde at en av gruppemedlemmene måtte følge med på PCen som hadde nedlastingen gående. Noe som resulterte i at nedlastinger flere ganger ble avbrutt, men ikke startet igjen før det ble lagt merke til.

[1] `https://metrics.torproject.org/onionperf-throughput.html?start=2023-04-12&end=2023-05-10&server=onion`

Problemer ved søkeroboten ikke får lastet ned filer

I testfasen ble det observert at søkeroboten ikke fikk lastet ned filer uten menneskelig interaksjon. Søkeroboten får trykket på nedlastingsknapper, men får ikke interagert med forgrunnsvinduet som nedlastingen åpner. På grunn av dette må en av gruppemedlemmene interagere med søkeroboten og godkjenne nedlastingen av filen. Løsningen av dette problemet, ble å prøve å laste ned filene manuelt. Disse dataene ble presentert i avsnitt 4.3.1. En potensiell annen løsning på problemet kan være å benytte Selenium sin funksjonalitet for å interagere med forgrunnsvinduer. Årsaken til at dette ikke var løsningen som ble brukt var at funksjonaliteten ikke ble oppdaget før utviklingen av søkeroboten var ferdig.

5.5 VPN på Tor

Når en bruker VPN over Tor skjules IP-adressen til brukeren for inngangsnoden. Trafikk som sendes gjennom Tor-nettverket krypteres hele veien, dermed vil ikke utgangsnoden kunne se hvilken data som sendes til nettsiden. Hvis Tor sine sikkerhetsmekanismer svikter, vil en VPN fungere som et ekstra sikkerhetslag. Dette vil gjøre at brukeren fortsatt kan holde seg anonym.

Ved bruk av en upålitelig VPN-tjeneste, kan tjenesten selge data videre, som kan brukes for å profilere brukeren. Videre vil heller ikke være en garanti for at trafikken som sendes over vil være krypert med en sikker nok krypteringsnøkkel. Uten en god nok kryptering vil ondsinnede aktører kunne dekryptere trafikken som overføres, og bruk av VPN vil miste all hensikt. Enkelte VPN-tjenester har dårlige serverløsninger, dette kan føre til lave nettverkshastigheter og flere nettverksbrudd.

Som nevnt i avsnitt 1.7 er et av resultatmålene å gjennomføre bacheloroppgaven uten at gruppemedlemmene skal bli eksponert på det mørke nettet, og et av læringsmålene er å opparbeide seg kunnskap om hvordan en kan holde seg sikker og anonym på det mørke nettet. Dette ble gjort ved bruk av VPN, fra tjenesten ExpressVPN, en godt anerkjent VPN-leverandør, som har gode rutiner rundt personvern og sikkerhet.

5.6 Erfaringer

Denne oppgaven tar for seg et svært omfattende tema, som gruppe hadde begrenset kunnskap om fra tidligere. Arbeidet med bacheloroppgaven bærer preg av en vag oppgavebeskrivelse. Oppdragsgivers formulering tydeliggjorde ikke forventningene til sluttresultatet, noe som førte til usikkerhet i gruppen. Oppgaven hadde fra start en viss retning, men ulike føringer underveis, på hvordan oppgaven skulle utføres og hva som skulle svares på, førte til mye unødvendig tidsbruk. Det ble

utforsket flere problemstillinger før en passende problemstilling ble identifisert til slutt. Det ble også forsøkt ulike metoder som en del av utviklingsarbeidet til søkeroboten før den mest optimale metoden ble oppdaget.

Til tross for alle utfordringene gruppen møtte på underveis, ble alle kravene til oppdragsgiver oppfylt og alle egendefinerte mål ble nådd. Disse vanskelighetene har bidratt til enda dypere forståelse av temaet samt gjort resultatet enda mer sammensatt. Formålet med oppgaven ble omdefinert underveis, men sluttresultatet anses som meget godt. Gjennom arbeidet med oppgavene har gruppemedlemmene lært svært mye om Tor og fått et mer nyansert bilde av teknologien.

Språkvalget som ble gjort resulterte i flere krevende situasjoner, og dette førte til at gruppen tidlig innså at det kunne ha vært valgt annerledes. I IT-fagfeltet er det et stort antall engelske ord og uttrykk, og det har ikke alltid vært enkelt å finne tilsvarende norske ord og uttrykk av samme kvalitet. Derfor opplevde gruppen at det ville vært enklere å skrive hele oppgaven på engelsk, ettersom de faglige begrepene på norsk ikke alltid oppnår samme presisjon som på engelsk.

I prosjektplanen ble det opprettet et Gantt-skjema, som brukes som en tidsplan for hele prosjektet. Figur 5.3 under viser et utsnitt av dette Gantt-skjemaet, men det ble, som det ofte blir, endringer underveis.

OPPGAVE	FREMGANG	START	SLUTT
Fase 1: Prosjektplan			
1.1 Prosjektplan	100 %	11.jan	27.jan
Fase 2: Informasjonshenting			
2.1 Oppsett av PC'er	100 %	11.jan	12.jan
2.2 Bakgrunn og historie av Dark Web	100 %	27.jan	20.feb
2.3 Teknisk oppbygging av Dark Web	100 %	27.jan	20.feb
2.4 Navigasjon inne i Dark Web	45 %	27.jan	03.apr
Fase 3: Design av crawler			
3.1 Informasjonshenting (eksisterende crawler-koder)	100 %	30.jan	24.feb
3.2 Oppsett og programmeringsspråk	100 %	01.feb	14.feb
Fase 4: Crawler			
4.1 Utvikling av crawler	70 %	06.feb	31.mar
4.2 Utvikling av database med API	50 %	06.feb	31.mar
4.3 Testing av crawler	30 %	20.feb	31.mar
Fase 5: Kjøring og høsting			
5.1 Kjøre crawleren	0 %	20.mar	10.mai
Fase 6: Rapportskriving			
6.1 Lage utkast av rapport til veileder	7 %	20.feb	03.apr
6.2 Kun skrive rapport	0 %	10.apr	10.mai
6.3 Finpuss og ferdigstilling av rapport	0 %	10.mai	16.mai
Fase 7: Overlevering av samlet data			
7.1 Overlevering av database med samlet data	0 %	22.mai	23.mai

Figur 5.3: Gantt-skjema i prosjektplanen

Når en sammenligner utsnittene i figur 5.3 og figur 5.4, kan en se at flere av prosessene som informasjonshenting og utvikling av søkerobot varte mye lengre enn opprinnelig tenkt. Informasjonshentingen ble ferdig 01.mai. Grunnen til at det tok lengre tid er fordi oppgavens mål og avgrensing ble endret en del i løpet av prosessen. Likevel hadde gruppen faste møter med oppdragsgiver hver uke

og jobbet jevnt og mye fra start. Dette gjorde at det var fremdrift, selv om fristdatoene ble forskjøvet. Førsteutkast ble levert til veileder 10.april, i tillegg et 2. utkast 26.april og et siste utkast 14.mai. Oppdragsgiver var i løpet av prosessen også behjelpelig med å lese oppgaven, slik at gruppen skulle få tilbakemeldinger underveis. I tillegg, for at oppdragsgiver skulle være inkludert siden dette er noe han skal bruke videre i sin forskning. Oppgaven ble ferdigstilt i form av innhold 19.mai, slik at gruppen hadde helgen 20. og 21. til finpuss før leveringsfrist 22.mai.

OPPGAVE	FREMGANG	START	SLUTT
Fase 1: Prosjektplan			
1.1 Prosjektplan	100 %	11.jan	27.jan
Fase 2: Informasjonshenting			
2.1 Oppsett av PC'er	100 %	11.jan	12.jan
2.2 Bakgrunn og historie av Dark Web	100 %	27.jan	01.mai
2.3 Teknisk oppbygging av Dark Web	100 %	27.jan	01.mai
2.4 Navigasjon inne i Dark Web	100 %	27.jan	03.apr
Fase 3: Design av crawler			
3.1 Informasjonshenting (eksisterende crawler-koder)	100 %	30.jan	24.feb
3.2 Oppsett og programmeringsspråk	100 %	01.feb	14.feb
Fase 4: Crawler			
4.1 Utvikling av crawler	100 %	06.feb	12.apr
4.2 Utvikling av database med API	0 %	06.feb	31.mar
4.3 Testing av crawler	100 %	20.feb	12.apr
Fase 5: Kjøring og høsting			
5.1 Kjøre crawleren	100 %	12.apr	10.mai
Fase 6: Rapportskriving			
6.1 Lage utkast av rapport til veileder	100 %	20.feb	10.apr
6.2 Kun skrive rapport	100 %	10.apr	19.mai
6.3 Finpuss og ferdigstilling av rapport	100 %	19.mai	21.mai
Fase 7: Overlevering av samlet data			
7.1 Overlevering av database med samlet data	0 %	22.mai	23.mai

Figur 5.4: Oppdatert Gantt-skjema

5.7 Refleksjon

I henhold til diskusjonen presentert i avsnitt 5.6, ble det ikke fastsatt klare begrensninger med hensyn til omfanget av oppgaven. Dette resulterte i at omfanget av oppgaven vokste betydelig utover det som opprinnelig var planlagt, og forventet av oppdragsgiver. Det ble benyttet tre ulike metoder for å besvare problemstillingen, noe som i ettertid viste seg å være noe i overkant. De tre forskjellige metodene krevde hver sin tilnærming, som har vært utfordrende for alle gruppemedlemmene. I samråd med oppdragsgiver ble det gjennomført en litteraturstudie for å oppnå nødvendig bakgrunnskunnskap i forkant av utviklingen av søkeroboten som skulle kjøres på Tor. Gruppen manglet kunnskap knyttet til utførelsen av en litteraturstudie, dette måtte derfor tilegnes underveis i prosessen. Opprinnelig var intensjonen å begrense litteraturstudien til forskningsartikler relatert til temaet fordeler med Tor. Imidlertid ble det mot slutten av skriveprosessen antydet at også informasjonsinnhentingen gjennom internettressurser skulle inkluderes i litteraturstudien. Ettersom denne informasjonen ikke ble gitt tidligere, valgte

gruppen å beholde den eksisterende strukturen.

Det ble også utført en spørreundersøkelse for å besvare en del av problemstillingen. Resultatene fra denne spørreundersøkelsen blir både presentert og diskutert i denne oppgaven. I etterkant av utførelsen har gruppen konkludert med at spørreundersøkelsen kunne vært laget på engelsk slik at den kunne blitt delt på enda flere medier, som mest sannsynlig hadde økt antall besvarelser. Analyseringen av resultatetene fra spørreundersøkelsen antydet til at spørsmålene som ble stilt kan ha vært ledende. Gruppen innså også at enkelte av spørsmålene kanskje var litt vage og derfor kan ha blitt tolket feil av deltakerne. Det ble konkludert med at deltakerne burde hatt muligheten til å oppgi svarene sine i tekstformat, på spesifikke spørsmål, slik at det var mulig å velge noe annet enn det som ble presentert som svaralternativer. Baktanken bak svaralternativer var å optimalisere spørreundersøkelsen slik at den var kort og enkel. Noe som ble sett på som en nødvendighet for å få flest mulig besvarelser. Sett i etterkant kan dette ha påvirket resultatet negativt ettersom deltakerne ble tvunget til å besvare spørreundersøkelsen, kun ved hjelp av forhåndsbestemte svaralternativer. Det var kun et spørsmål som lot deltakerne sende inn svar i form av fritekst, og dette ga gruppen mer innsikt.

For å forsøke og høste mest mulig data som senere kan bli brukt av oppdragsgiver, ble det utviklet en søkerobot. I etterkant av utviklingen har det blitt observert enkelte forbedringspotensialer, som kunne gjort skrapeprosessen mer effektiv og oversiktlig. En av forbedringene vil som nevnt i avsnitt 5.2, være å ekskludere alle V2 onion-lenker. For en mer effektiv analyse av datamengder, ville det vært hensiktsmessig å implementere forbedrede logging-systemer for søkeroboten. Data som burde vært inkludert i dette logging-systemet er hvor mange av onion-lenkene som hadde blitt prøvd besøkt, men ikke lengre var operative. Dette ville hjulpet med å gi en oversikt over hvor stor andel av tiden søkeroboten brukte på nettsider som ikke lengre fungerte. På det mørke nettet er det enkelte nettsteder som bruker speilede nettsider, altså en kopi av sin egen nettside, men med en annen URL. Figur 3.6 viser hvordan nettstedet til Lockbit har lenker til 9 speilede sider liggende i DDoS-Protection sin. Søkeroboten kunne blitt programmert til å sjekke om den har funnet en lik side tidligere, for å ikke lagre duplikater av nettsteder.

5.8 Videre arbeid

Oppdragsgivers forskning

Oppdragsgiver skal, som nevnt tidligere, bruke informasjonen og dataen innhentet i denne rapporten i sin doktorgrad, der han forsker på hvordan avanserte trusselaktører opererer. Derfor er dataen gruppen har høstet interessant, samt at det er mulig å videreutvikle søkerobot for å hente enda mer data. Selv om det faktisk ble høstet en del data fra Tor, så er det absolutt mulig å få tak i mer. Søkerobo-

ten bør kanskje kjøre over en enda lengre periode, for eksempel noen måneder, for å få hentet ut større mengder ønsket data. Videre å tillegge søkeroboten flere funksjoner som vil spesifisere søket dens mer, som at den kan søke etter eksakte nøkkelord for å effektivisere høstingen. I tillegg funksjoner som gjør søkeroboten fullt ut automatisert og at den kan kjøre på egenhånd uten at en også må laste ned data manuelt. Dessuten funksjoner som at søkeroboten kan omgå sikkerhetsmekanismene gruppen møtte på i eget forsøk. Oppretting av brukere, CAPTCHA, beskyttelse mot tjenestenektangrep og lignende. Det vil også være mulig å gjøre søkeroboten om til et flertrådet program, slik at flere søkeroboter kan kjøres fra samme maskin samtidig. Det vil være interessant å se om høstede data vil tilføye noe i oppdragsgivers videre forskning.

Flere målinger og innhenting av data

Litteraturstudien kan gjøres mer omfattende, der det inkluderes flere forskningsartikler som belyser tema. Det er også mulig å utvide søket til flere databaser, som kan gi flere mulige resultater.

Når det gjelder spørreundersøkelsen kan den gjøres om til engelsk for å nå et bredere publikum. Eventuelt at en lager en både på norsk og på engelsk, da er det viktig at spørsmålene har helt samme betydning, slik at resultatene kan slås sammen.

Videre vil det kunne gjøres en større analyse av dataen som er høstet. For eksempel kan det undersøkes nærmere og sorteres ut fra type aktivitet, hvilke typer trusselaktører det gjelder og hvem som har blitt utsatt for angrep. I tillegg datoer for når de ulike dataene har blitt lastet ned, ordtelling, hvilket språk dataene er på og andre metadata.

Kapittel 6

Konklusjon

Problemformuleringen beskriver hva som var ønskelig å oppnå med denne oppgaven. I dette bachelorprosjektet gjaldt det å først innhente informasjon om Tor for å lære mest mulig om bakgrunn, hensikt og teknisk oppbygging. Etter å ha fullført denne oppgaven har gruppen oppnådd betydelig innsikt i Tor og besitter en omfattende kunnskapsbase, selv om det er både mulig og sannsynlig at det fremdeles finnes mer informasjon tilgjengelig.

Neste punkt var å utføre en utforskning av fordeler med Tor. Parallelt med dette ble det ble det utført en spørreundersøkelse med det formål å undersøke befolkningens oppfattelse av det mørke nettet. Gruppen, som mange andre, hadde negative holdninger når det gjaldt Tor og det mørke nettet generelt. De negative holdningene til befolkningen kom frem både i gruppens egen spørreundersøkelse, i tillegg til i undersøkelsen utført av CIGI. Likevel, etter å ha utført dette dypdykket i Tor, er det nå etablert kunnskap om dets mangfoldige anvendelser som går utover kriminell aktivitet. Gruppen vet nå at Tor ikke utelukkende er preget av negative aspekter.

Å utvikle en søkerobot for å høste data var det siste punktet. På bakgrunn av at Tor er bygd opp som det er, måtte søkeroboten utvikles på en annen måte enn om det var det åpne nettet gruppen skulle hente data fra. Dette gjorde oppgaven mer krevende, men gruppen løste likevel dette. Teknisk sett har Tor en høy brukerterskel, det er ikke intuitivt å finne ting der. Oppdragsgiver hadde klare mål om hvilke data han ønsket å få tilbake etter endt prosjekt, noe gruppen nå kan levere.

Videre hadde gruppen en rekke prosjektmål som skulle oppnås. Effektmål, resultatmål og læringsmål:

- **E1**: Få oversikt over oppbyggingen, bakgrunnen og historien til Tor, samt data som ligger der.
- **E2**: Utvikle en søkerobot som høster dataene nevnt over.
- **E3**: Få hentet ut mest mulig informasjon om hackergrupper fra det mørke nettet via Tor, slik at det kan forskes videre på.

- **R1**: Ha en ferdigutviklet søkerobot som kan hente ned innhold fra det mørke nettet relatert til nøkkelord en vil lete etter.
- **R2**: Å ha en leveringsklar og oversiktlig database med ulike data fra det mørke nettet.
- **R3**: Gjennomføre bacheloroppgaven på en sikker måte, slik at ingen på gruppen blir eksponert på det mørke nettet.

- **L1**: Opparbeide kunnskap om det mørke nettet, Tor, hvordan det fungerer og henger sammen.
- **L2**: Opparbeide kunnskap om hvordan en kan holde seg sikker og anonym på det mørke nettet.
- **L3**: Gjennomføre et større prosjektarbeid og metodisk arbeid.
- **L4**: Hvordan lage en søkerobot og hvordan den fungerer.

Effektmål 1 og 2 kan vi konkludere som oppnådd. Effektmål 3 er også oppnådd, selv om "mest mulig" er en subjektiv tolkning. Gruppen hentet den dataen som var mulig å få tak i på den tiden som ble satt.

Resultatmål 1 er delvis oppnådd, gruppen valgte etter hvert å ikke forsøke å implementere funksjonen for å legge inn nøkkelord for søkeroboten. Ettersom det var svært vanskelig å finne alle nøkkelord som hadde gitt merverdi til høstingen, ville en slik funksjon mest sannsynlig da ført til at flere relevante nettsider hadde blitt oversett. For at søkeroboten skulle finne mest mulig relevant data som kunne brukes av oppdragsgiver, ble heller fokuset på å finne URLer knyttet til denne dataen.

Resultatmål 2 og 3 er realisert. Læringsmålene er knyttet opp mot effektmål og resultatmål, og gruppen sitter igjen med veldig mye ny kunnskap etter gjennomført bacheloroppgave. Alle læringsmålene er oppnådd.

Tilslutt problemstillingen med at det ikke finnes en enkelt kilde som inkluderer en helhetlig og bred forståelse av Tor på ett og samme sted. Det kan konkluderes med at dette er noe gruppen har fått til med denne rapporten. I tillegg har fordelene ved Tor blitt belyst omgående, og gruppen mener det har tydeliggjort at det ikke kun er negative sider angående Tors bruksområder.

Bibliografi

[1] Norsis. «Hva er løsepengeangrep.» (2. aug. 2022), adresse: https://norsis.no/fakta/hva-er-losepengeangrep/ (sjekket 28.03.2023).

[2] digi. «Identiteten til tusenvis av nordmenn selges på det mørke nettet.» (13. des. 2022), adresse: https://www.digi.no/artikler/identiteten-til-tusenvis-av-nordmenn-selges-pa-det-morke-nettet/524243 (sjekket 28.03.2023).

[3] «Kva er Digitaliseringsdirektoratet?» Digitaliseringsdirektoratet. (18. apr. 2023), adresse: https://www.digdir.no/digdir/kva-er-digitaliseringsdirektoratet/703 (sjekket 18.04.2023).

[4] «ENISA Threat Landcape 2022,» s. 22–23, 3. nov. 2022. adresse: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2022 (sjekket 16.01.2023).

[5] CrowdStrike, «EXPOSING THE OPEN, DEEP, AND DARK WEB AND BEYOND,» s. 3, 2021.

[6] «History,» Tor Project. (4. mar. 2023), adresse: https://www.torproject.org/about/history/ (sjekket 15.03.2023).

[7] A. K. Ghazi-Tehrani, «Mapping Real-World Use of the Onion Router,» s. 245, 2023.

[8] G. O. E. Jardine A. M. Lindner, «The potential harms of the Tor anonymity network cluster disproportionately in free countries,» s. 1–6, 2020.

[9] H. Øverby og H. Dvergsdal. «internett,» Store norsk leksikon. (), adresse: https://snl.no/internett (sjekket 13.04.2023).

[10] «Hva er det dype og mørke web,» Kaspersky. (), adresse: https://www.kaspersky.no/resource-center/threats/deep-web (sjekket 06.04.2023).

[11] A. Volle. «Dark Web,» Britannica. (14. mar. 2023), adresse: https://www.britannica.com/topic/dark-web (sjekket 12.04.2023).

[12] «TOR network,» Wikipedia. (feb. 2023), adresse: https://en.wikipedia.org/wiki/Tor_(network)#History (sjekket 16.02.2023).

[13] «12 Legitimate Uses for the Dark Web,» Make Use Of. (10. mai 2022), adresse: https://www.makeuseof.com/dark-web-legitimate-uses/ (sjekket 02.05.2023).

[14] S. Retzkin, *Hands-On Dark Web Analysis: Learn what goes on in the Dark Web, and how to work with it*. Packt Publishing, 2018, ISBN: 178913336X.

[15] «The Guardian SecureDrop,» The Guardian. (), adresse: https://www.theguardian.com/securedrop (sjekket 24.04.2023).

[16] «Den arabiske våren,» Store norske leksikon. (21. des. 2022), adresse: https://snl.no/Den_arabiske_v%5C%C3%5C%A5ren (sjekket 06.04.2023).

[17] «Arab Spring anniversary: When Egypt cut the internet,» Al Jazeera. (25. jan. 2016), adresse: https://www.aljazeera.com/features/2016/1/25/arab-spring-anniversary-when-egypt-cut-the-internet (sjekket 17.04.2023).

[18] «Internet censorship in the Arab Spring,» Wikipedia. (3. apr. 2023), adresse: https://en.wikipedia.org/wiki/Internet_censorship_in_the_Arab_Spring (sjekket 06.04.2023).

[19] «Velkommen til Tors verden,» NRK. (3. okt. 2013), adresse: https://nrkbeta.no/2013/10/03/velkommen-til-tors-verden/ (sjekket 06.04.2023).

[20] «Federal Service for Supervision of Communications, Information Technology and Mass Media,» The Russian Government. (), adresse: http://government.ru/en/department/58/ (sjekket 26.04.2023).

[21] T. Project, «ANNUAL REPORT 2020-2021,» 2021, s. 4–6.

[22] «Snowflake,» Tor Project. (), adresse: https://snowflake.torproject.org (sjekket 26.04.2023).

[23] «Is Tor browser safe to use?» NordVPN. (7. jan. 2023), adresse: https://nordvpn.com/no/blog/is-tor-safe/ (sjekket 17.04.2023).

[24] M. Huang Hsiao-Ying Bashir, «Proceedings of the ASIST Annual Meeting: The onion router. Understanding a privacy enhancing technology community,» 2016, s. 1–10.

[25] G. Greenwald, «NSA collecting phone records of millions of Verizon customers daily,» *The Guardian*, 6. jun. 2013. adresse: https://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order (sjekket 12.04.2023).

[26] L. Poitras og G. Greenwald. «NSA whistleblower Edward Snowden: 'I don't want to live in a society that does these sort of things' – video,» The Guardian. (), adresse: https://www.theguardian.com/world/video/2013/jun/09/nsa-whistleblower-edward-snowden-interview-video.

[27] G. Greenwald, «XKeyscore: NSA tool collects 'nearly everything a user does on the internet',» *The Guardian*, adresse: https://www.theguardian.com/world/2013/jul/31/nsa-top-secret-program-online-datar (sjekket 12.04.2023).

[28] H. Carlsen, «Overvåkingsprogram overvåker nesten alt du gjør på nettet,» *NRK*, adresse: https://www.nrk.no/urix/_-nsa-kan-overvake-alt-du-gjor-1.11159299 (sjekket 13.04.2023).

[29] «The Bright Side of the Dark Web,» DarkReading. (15. jun. 2020), adresse: https://www.darkreading.com/risk/the-bright-side-of-the-dark-web (sjekket 12.04.2023).

[30] C. for International Governance Innovation. «About CIGI.» (), adresse: https://www.cigionline.org/about/ (sjekket 10.05.2023).

[31] C. for International Governance Innovation. «CIGI-IPSOS GLOBAL SURVEY INTERNET SECURITY TRUST.» (2019), adresse: https://www.cigionline.org/sites/default/files/documents/2019%5C%20CIGI-Ipsos%5C%20Global%5C%20Survey%5C%20-%5C%20Part%5C%205%5C%20Cryptocurrencies%5C%2C%5C%20Blockchain%5C%2C%5C%20Dark%5C%20Web%5C%20%5C%26%5C%20Product%5C%20Certification.pdf (sjekket 11.05.2023).

[32] «What are Entry Guards?» the Tor Project. (3. apr. 2023), adresse: https://support.torproject.org/about/entry-guards/ (sjekket 18.04.2023).

[33] «Tor 0.4.8.0-alpha-dev,» Tor Project. (), adresse: https://tpo.pages.torproject.net/core/doc/tor/entrynodes_8c_source.html (sjekket 16.04.2023).

[34] «Onion Routing,» GeeksforGeeks. (17. mar. 2022), adresse: https://www.geeksforgeeks.org/onion-routing/ (sjekket 24.02.2023).

[35] «Types Of Relays On The Tor Network,» Tor project. (13. apr. 2023), adresse: https://community.torproject.org/relay/types-of-relays/ (sjekket 15.04.2023).

[36] «The New York Times is Now Available as a Tor Onion Service,» NYT Open. (12. feb. 2022), adresse: https://open.nytimes.com/https-open-nytimes-com-the-new-york-times-as-a-tor-onion-service-e0d0b67b7482 (sjekket 03.05.2023).

[37] R. W. Håkon Bergsjø og L. Øverlier, *Digital Sikkerhet: en innføring*. Universitetsforlaget, 2020, ISBN: 9788215034225.

[38] «.onion,» Wikipedia. (26. mar. 2023), adresse: https://en.wikipedia.org/wiki/.onion (sjekket 06.04.2023).

[39] «løk-ruting,» SNL. (5. aug. 2022), adresse: https://snl.no/l%C3%B8k-ruting (sjekket 24.02.2023).

[40] «Really Private Browsing: An Unofficial User's Guide to Tor,» Make Use Of. (31. jul. 2017), adresse: https://www.makeuseof.com/tag/really-private-browsing-an-unofficial-users-guide-to-tor/ (sjekket 02.03.2023).

[41] «Søkerobotene gjør deg en tjeneste,» Idium. (), adresse: https://www.idium.no/b/soekerobotene-gjoer-deg-en-tjeneste (sjekket 26.04.2023).

[42] «25 Best Web Crawler Tools,» Startup Stash. (6. des. 2022), adresse: https://startupstash.com/web-crawler-tools/ (sjekket 03.05.2023).

[43] «parsing,» Store norsk leksikon. (15. feb. 2023), adresse: https://snl.no/parsing (sjekket 07.04.2023).

[44] «Beautiful Soup (HTML parser),» Wikipedia. (27. mar. 2023), adresse: https://en.wikipedia.org/wiki/Beautiful_Soup_(HTML_parser) (sjekket 07.04.2023).

[45] «Python Requests Tutorial,» GeeksforGeeks. (8. feb. 2023), adresse: https://www.geeksforgeeks.org/python-requests-tutorial/ (sjekket 07.04.2023).

[46] «The Selenium Browser Automation Project,» Selenium. (2. mar. 2023), adresse: https://www.selenium.dev/documentation/ (sjekket 07.04.2023).

[47] «Nasjonalt cyberkrimsenter (NC3),» Politiet. (), adresse: https://www.politiet.no/om-politiet/organisasjonen/sarorganene/kripos/kripos-hovedarbeidsomrader/nasjonalt-cyberkrimsenter/ (sjekket 18.04.2023).

[48] H. Aveyard, *Doing a Literature Review in Health and Social Care : A Practical Guide (2nd Edition)*. Open University Press, 2014, ISBN: 9780335263073.

[49] «Hsiao-Ying Huang,» Google Scholar. (), adresse: https://scholar.google.com/citations?user=2iTIIVEAAAAJ&hl=en (sjekket 17.05.2023).

[50] «Masooda Bashir,» University of Illinois. (), adresse: https://ischool.illinois.edu/people/masooda-bashir (sjekket 17.05.2023).

[51] «Eric Jardine,» Google Scholar. (), adresse: https://scholar.google.ca/citations?user=WC3tk7kAAAAJ&hl=en (sjekket 17.05.2023).

[52] «Andrew M. Lindner - ASSOCIATE PROFESSOR OF SOCIOLOGY,» Skidmore Univeristy. (), adresse: https://www.skidmore.edu/sociology/faculty/lindner.php (sjekket 17.05.2023).

[53] «Gareth Owenson,» Linkedin. (), adresse: https://www.linkedin.com/in/gareth-owenson-a89aa5202/?originalSubdomain=uk (sjekket 17.05.2023).

[54] «Company,» Searchlight Cyber. (), adresse: https://www.slcyber.io/company/ (sjekket 17.05.2023).

[55] «Google Forms,» Google. (mar. 2023), adresse: https://www.google.com/forms/about/ (sjekket 09.03.2023).

[56] «Python Package Index,» Python Software Foundation. (mar. 2023), adresse: https://pypi.org (sjekket 10.03.2023).

[57] «5 Preferred Programming Languages for Web Scraping,» Open Data Science. (11. aug. 2022), adresse: https://opendatascience.com/5-preferred-programming-languages-for-web-scraping/ (sjekket 03.05.2023).

[58] «Documentation for Visual Studio Code,» Microsoft. (), adresse: https://code.visualstudio.com/docs (sjekket 03.03.2023).

[59] «Learn PyCharm,» JetBrains. (2. mar. 2023), adresse: https://www.jetbrains.com/pycharm/learn/ (sjekket 06.03.2023).

[60] «Linux Lite Requirements,» Linux Lite. (), adresse: https://www.linuxliteos.com/download.php#requirements (sjekket 28.04.2023).

[61] «Writing the Linux Lite ISO to a USB on Windows,» Linux Lite. (), adresse: https://www.linuxliteos.com/manual/install.html#llusbwin (sjekket 28.04.2023).

[62] P. Batard. «Rufus.» (26. apr. 2023), adresse: https://rufus.ie/nb/ (sjekket 28.04.2023).

[63] «THE MAIN LINUX FILE SYSTEMS AND THEIR DIFFERENCES,» VEPROF. (12. aug. 2021), adresse: https://www.veprof.com/blog/technology/linux-file-systems-and-their-differences (sjekket 28.04.2023).

[64] «De beste VPN tjenestene for Norge 2023,» ITavisen. (), adresse: https://itavisen.no/vpn/ (sjekket 16.05.2023).

[65] «ExpressVPN Privacy Policy,» ExpressVPN. (), adresse: https://www.expressvpn.com/no/privacy-policy (sjekket 16.05.2023).

[66] T. H. Nätt. «skraping (IT).» (23. feb. 2022), adresse: https://snl.no/skraping_-_IT (sjekket 05.05.2023).

[67] C. Dilmegani. «Web Crawler: What It Is, How It Works Applications in 2023.» (28. mar. 2023), adresse: https://research.aimultiple.com/web-crawler/ (sjekket 06.05.2023).

[68] Selenium. «Working with windows and tabs.» (20. sep. 2022), adresse: https://www.selenium.dev/documentation/webdriver/interactions/windows/#takescreenshot (sjekket 11.05.2023).

[69] «DEEP WEB VS DARK WEB: WHAT'S THE DIFFERENCE?» Crowdstrike. (25. okt. 2022), adresse: https://www.crowdstrike.com/cybersecurity-101/the-dark-web-explained/deep-web-vs-dark-web/ (sjekket 03.05.2023).

[70] «Difference between Internal and External fragmentation,» GeeksForGeeks. (21. feb. 2023), adresse: https://www.geeksforgeeks.org/difference-between-internal-and-external-fragmentation/ (sjekket 15.05.2023).

[71] «Tor is slow right now. Here is what is happening.,» Tor project. (7. feb. 2020), adresse: https://blog.torproject.org/tor-network-ddos-attack/ (sjekket 23.03.2023).

[72] «Tor Metrics: Performance,» Tor Project. (), adresse: https://metrics.torproject.org/onionperf-throughput.html?start=2023-04-12&end=2023-05-10&server=onion (sjekket 10.05.2023).

Akronymer

CIGI Centre for International Governance Innovation. 16, 18, 64, 66, 67, 79

DoS Denial of Service. 14

EFF Electronic Frontier Foundation. 11

ENISA The European Union Agency for Cybersecurity. 1

exFAT Extensible File Allocation Table. 39, 40

HTTP Hypertext Transfer Protocol. 71

IT Informasjonsteknologi. 37, 75

Mbps Megabit per sekund. 72

NC3 Nasjonalt cyberkrimsenter. 27, 41, 68

NRL Naval Research Laboratory. 11

PET Privacy Enhancing Technologies. 15

PyPI Python Package Index. 38

Tor The Onion Router. 2, 3, 4

VSC Visual Studio Code. 39

Ordliste

alfaversjon Alfaversjon er en tidlig utgave av et program eller en applikasjon, ofte ikke helt stabil, men kan vise hva programmet kan gjøre. Som en forhåndsvisning [1]. 11

CAPTCHA Completely Automated Public Turing test to tell Computers and Humans Apart. Sikkerhetsmekanisme for å verifisere at en handling blir utført av et menneske og ikke en maskin. Hensikten er å forhindre angrep der programvare benyttes til å automatisere handlinger slik som registrering av brukerkontoer eller forespørsler om datauthenting og dataprosessering. [2]. xiv, 27, 44, 46, 48, 61, 62, 69, 70, 72, 78

CSS CSS, format for stilsett for html-dokumenter på internett, standardisert av W3C. Stilsett er en enkel måte å definere stiler på, det vil si layoutegenskaper som skrifttyper, farger og linjeavstander, slik at alle dokumenter fra et gitt nettsted får samme design.[3]. 60, 62

datakriminalitet Datakriminalitet er straffbare handlinger som forutsetter utnyttelse av datateknologi. Slike handlinger kan grovt sett deles i tre undergrupper: endring og sletting av data, urettmessig innsyn i og bruk av data (tapping), og ulovlig bruk av datautstyr.[4]. 3

dekryptering Dekryptering er det å omforme krypterte data slik at de igjen blir lesbare. Dette gjøres ved bruk av en krypteringsalgoritme og en krypteringsnøkkel [5]. 22

DNS DNS er en grunnleggende internett-tjeneste levert av spesielle DNS-servere (også kalt «rotservere») som gjør at andre internett-servere finner frem til hverandre [6]. 24

GET GET er en metode brukt til å forespørre data fra en spesifisert kilde[7]. 23

[1] https://en.wiktionary.org/wiki/alpha_version
[2] https://snl.no/CAPTCHA
[3] https://snl.no/CSS
[4] https://snl.no/datakriminalitet
[5] https://snl.no/dekryptering
[6] https://snl.no/DNS
[7] https://www.w3schools.com/tags/ref_httpmethods.asp

grafisk brukergrensesnitt Brukergrensesnitt er innen IT en betegnelse på kontaktflaten mellom brukeren og et system. Et grafisk brukergrensesnitt vil i de fleste tilfeller kontrolleres via berøringsskjerm eller mus, der man peker og klikker på knapper og ikoner for å styre programmer.[8]. 24

hopp Hopp spesifiserer et mål på antall enheter som må traverseres for å nå destinasjonsenheten.[9]. 20, 40

HTML HTML er et markeringsspråk (engelsk markup language) for å utvikle nettsider. I dag er det i praksis en enerådende standard for innholdet i nettsider, og HTML støttes av alle vanlige nettlesere.[10]. 24, 42, 44, 45, 58, 62

indeksert Indeksering er prosessen som ordner et sett med data slik at det blir mulig å finne dem basert på søkeverdi. Søkemotores indeks samler, setter sammen og lagrer data for rask og presis gjnefinning av informasjon. Oppbyggingen av indeksen bygger på lingvistikk, kognitiv psykologi, matematikk, informatikk, fysikk og datafag [11]

. 11

informasjonskapsel Informasjonskapsel, eller cookie, er en liten datamengde som en nettleser mottar fra nettsider som brukeren besøker, og som lagres som en fil på brukerens harddisk [12]. 24

IP-adresse En IP-adresse er en unik adresse som identifiserer en enhet på Internett eller et lokalt nettverk. IP står for «Internet Protocol» (internettprotokoll), som er et sett med regler som styrer formatet på dataene som endes via Internett eller det lokale nettverket. [13]. 20, 26, 40, 74

kryptering Kryptering innebærer å omforme data slik at de ikke kan leses eller endres av noen som urettmessig får tilgang til de. Dette gjøres vanligvis ved bruk av en kjent krypteringsalgoritme og en hemmelig krypteringsnøkkel [14]. 22, 24

kø En kø er definert som en lineær datastruktur som er åpen i begge ender, og operasjonene er utført etter Først inn, Først ut (FIFO) prinsippet. [15]. 26

[8] https://snl.no/brukergrensesnitt

[9] https://documentation.solarwinds.com/en/success_center/ntm/content/ntm_what_are_hops.htm

[10] https://snl.no/HTML

[11] https://no.wikipedia.org/wiki/Indeksering_(datateknologi)

[12] https://snl.no/informasjonskapsel

[13] https://www.kaspersky.no/resource-center/definitions/what-is-an-ip-address

[14] https://snl.no/kryptering

[15] https://www.geeksforgeeks.org/queue-data-structure/

Linux Linux er en familie av operativsystemer. Per 2018 kjører de fleste av verdens servere, og nærmest alle superdatamaskiner Linux, men det finnes også populære varianter til personlige datamaskiner.[16]. 40

løsepengeangrep En type skadevare som krypterer eller låser hele eller deler av innholdet på datamaskinen. For å få tilgang igjen krever angriperne at man betaler løsepenger. [17]. 1

metadata Metadata er informasjon som beskriver annen informasjon, altså data om data. Innen IT vil metadata ofte være tilleggsinformasjon i elektroniske filformater slik som tekstdokumenter, bildefiler eller videofiler.[18]. 24

nettverkslaget Nettverkslaget skal sørge for at dataene kommer frem i riktig rekkefølge til riktig mottaker. Den mest kjente protokollen på dette nivået er IP (Internet Protocol).[19]. 21

noder I denne oppgaven betyr det TOR-noder som vil si spesialiserte servere i nettverket som formidler trafikk mellom en sender og en mottaker [20]. 11, 15

off. PhD Offentlig sektor ph.d. Der en arbeidsgiver i offentlig sektor kan søke støtte for at en ansatt kan kan gjennomføre et doktorgradsprosjekt [21]. 1, 30

proxy Proxy brukes også som forkortelse for proxyserver, en server som opererer som et mellomledd mellom en klient og en server (oftest på vegne av klienten), og som filtrerer, behandler, lagrer informasjon, vanligvis for bedre ytelse, hemmelighold eller sikkerhet [22]. 27

Raspberry Pi En Raspberry Pi er en liten datamaskin på størrelse med et bankkort. [23]. 40

stakk Stakk er en datastruktur hvor data hentes ut i motsatt rekkefølge av den de blir lagt inn i. Strukturen fungerer som en stabel hvor man kun kan legge til og fjerne elementer fra toppen, følger altså Sist inn, Først ut (LIFO) prinsippet. [24]. 26

---

[16] https://support.google.com/googlenest/answer/6274087?hl=no

[17] https://norsis.no/fakta/hva-er-losepengeangrep/

[18] https://snl.no/metadata

[19] https://snl.no/OSI_-_datakommunikasjon

[20] https://metadon.no/index.php/hvordan-fungerer-tor-nettleseren/

[21] https://www.forskningsradet.no/sok-om-finansiering/midler-fra-forskningsradet/offentlig-sektor-phd/

[22] https://snl.no/proxy

[23] https://www.spiceworks.com/tech/networking/articles/what-is-raspberry-pi/

[24] https://snl.no/stakk_-_IT

søkerobot Dataprogram, bot som systematisk crawler internett, vanligvis for å sanke og kategorisere informasjon; crawler [25]. v, vii, x, 3, 4, 5, 6, 9, 24, 25, 26, 29, 37, 38, 39, 40, 41, 44, 45, 46, 47, 48, 54, 55, 56, 61, 62, 63, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80

tjenestenektangrep DDoS - Distributed denial of service; Et angrep som hindrer at noe eller noen (et system eller en person) får tilgang til informasjon eller ressurser de vil ha tilgang til.[26]. 46, 71, 78

URL Uniform Resource Locator. URL er en standard for angivelse av adresser til dokumenter og andre ressurser. De fleste forbinder URL med ressurser på internett, og nettadresse eller web-adresse er andre vanlige navn. [27]. xiii, xiv, 21, 24, 25, 26, 40, 41, 42, 44, 45, 46, 48, 54, 55, 56, 58, 67, 70, 72, 73, 77, 80

VPN Virtuelt Privat Nettverk; Når private datamaskiner på fremmede steder kobler seg til krypterte forbindelser gjennom svitsjede nett til et annet nett.[28]. x, 3, 6, 14, 26, 27, 40, 47, 51, 74

[25] https://naob.no/ordbok/sÃÿkerobot
[26] https://nettvett.no/ddos-angrep/
[27] https://snl.no/URL
[28] https://snl.no/VPN

Vedlegg A

Prosjektavtale

NTNU

Norges teknisk-naturvitenskapelige universitet

Fastsatt av prorektor for utdanning 10.12.2020

STANDARDAVTALE

om utføring av studentoppgave i samarbeid med ekstern virksomhet

Avtalen er ufravikelig for studentoppgaver (heretter oppgave) ved NTNU som utføres i samarbeid med ekstern virksomhet.

Forklaring av begrep

Opphavsrett
Er den rett som den som skaper et åndsverk har til å fremstille eksemplar av åndsverket og gjøre det tilgjengelig for allmennheten. Et åndsverk kan være et litterært, vitenskapelig eller kunstnerisk verk. En studentoppgave vil være et åndsverk.

Eiendomsrett til resultater
Betyr at den som eier resultatene bestemmer over disse. Utgangspunktet er at studenten eier resultatene fra sitt studentarbeid. Studenten kan også overføre eiendomsretten til den eksterne virksomheten.

Bruksrett til resultater
Den som eier resultatene kan gi andre en rett til å bruke resultatene, f.eks. at studenten gir NTNU og den eksterne virksomheten rett til å bruke resultatene fra studentoppgaven i deres virksomhet.

Prosjektbakgrunn
Det partene i avtalen har med seg inn i prosjektet, dvs. som vedkommende eier eller har rettigheter til fra før og som brukes i det videre arbeidet med studentoppgaven. Dette kan også være materiale som tredjepersoner (som ikke er part i avtalen) har rettigheter til.

Utsatt offentliggjøring
Betyr at oppgaven ikke blir tilgjengelig for allmennheten før etter en viss tid, f.eks. før etter tre år. Da vil det kun være veileder ved NTNU, sensorene og den eksterne virksomheten som har tilgang til studentarbeidet de tre første årene etter at studentarbeidet er innlevert.

1. Avtaleparter

Norges teknisk-naturvitenskapelige universitet (NTNU) Institutt: Institutt for informasjonssikkerhet og kommunikasjonsteknologi
Veileder ved NTNU: Erjon Zoto e-post og tlf. erjon.zoto@ntnu.no, 98 43 30 97
Ekstern virksomhet: Digitaliseringsdirektoratet (DIGDIR) Ekstern virksomhet sin kontaktperson, e-post og tlf.: Raymond André Hagen, raymohag@stud.ntnu.no, 92 68 57 71
Student: Marte Jørgensen Fødselsdato: 23.01.1994
Student: Martin Hyldmo Fødselsdato: 01.04.01
Student: Mats Dimitri Jensen Fødselsdato: 05.05.97
Student: Thugitha Kanavathi Fødselsdato: 17.08.1999

Partene har ansvar for å klarere eventuelle immaterielle rettigheter som studenten, NTNU, den eksterne eller tredjeperson (som ikke er part i avtalen) har til prosjektbakgrunn før bruk i forbindelse med utførelse av oppgaven. Eierskap til prosjektbakgrunn skal fremgå av eget vedlegg til avtalen der dette kan ha betydning for utførelse av oppgaven.

2. Utførelse av oppgave

Studenten skal utføre: (sett kryss)

Masteroppgave	
Bacheloroppgave	x
Prosjektoppgave	
Annen oppgave	

Startdato: 09.01.23
Sluttdato: 22.05.23

Oppgavens arbeidstittel er: Høsting fra mørke nettet

Ansvarlig veileder ved NTNU har det overordnede faglige ansvaret for utforming og godkjenning av prosjektbeskrivelse og studentens læring.

3. Ekstern virksomhet sine plikter

Ekstern virksomhet skal stille med en kontaktperson som har nødvendig faglig kompetanse til å gi studenten tilstrekkelig veiledning i samarbeid med veileder ved NTNU. Ekstern kontaktperson fremgår i punkt 1.

Formålet med oppgaven er studentarbeid. Oppgaven utføres som ledd i studiet. Studenten skal ikke motta lønn eller lignende godtgjørelse fra den eksterne for studentarbeidet. Utgifter knyttet til gjennomføring av oppgaven skal dekkes av den eksterne. Aktuelle utgifter kan for eksempel være reiser, materialer for bygging av prototyp, innkjøp av prøver, tester på lab, kjemikalier. Studenten skal klarere dekning av utgifter med ekstern virksomhet på forhånd.

Ekstern virksomhet skal dekke følgende utgifter til utførelse av oppgaven:

Dekning av utgifter til annet enn det som er oppført her avgjøres av den eksterne underveis i arbeidet.

4. Studentens rettigheter

Studenten har opphavsrett til oppgaven[1]. Alle resultater av oppgaven, skapt av studenten alene gjennom arbeidet med oppgaven, eies av studenten med de begrensninger som følger av punkt 5, 6 og 7 nedenfor. Eiendomsretten til resultatene overføres til ekstern virksomhet hvis punkt 5 b er avkrysset eller for tilfelle som i punkt 6 (overføring ved patenterbare oppfinnelser).

I henhold til lov om opphavsrett til åndsverk beholder alltid studenten de ideelle rettigheter til eget åndsverk, dvs. retten til navngivelse og vern mot krenkende bruk.

Studenten har rett til å inngå egen avtale med NTNU om publisering av sin oppgave i NTNUs institusjonelle arkiv på Internett (NTNU Open). Studenten har også rett til å publisere oppgaven eller deler av den i andre sammenhenger dersom det ikke i denne avtalen er avtalt begrensninger i adgangen til å publisere, jf. punkt 8.

5. Den eksterne virksomheten sine rettigheter

Der oppgaven bygger på, eller videreutvikler materiale og/eller metoder (prosjektbakgrunn) som eies av den eksterne, eies prosjektbakgrunnen fortsatt av den eksterne. Hvis studenten skal utnytte resultater som inkluderer den eksterne sin prosjektbakgrunn, forutsetter dette at det er inngått egen avtale om dette mellom studenten og den eksterne virksomheten.

Alternativ a) (sett kryss) Hovedregel

[1] Jf. Lov om opphavsrett til åndsverk mv. av 15.06.2018 § 1

x	Ekstern virksomhet skal ha bruksrett til resultatene av oppgaven

Dette innebærer at ekstern virksomhet skal ha rett til å benytte resultatene av oppgaven i egen virksomhet. Retten er ikke-eksklusiv.

Alternativ b) (sett kryss) Unntak

	Ekstern virksomhet skal ha eiendomsretten til resultatene av oppgaven og studentens bidrag i ekstern virksomhet sitt prosjekt

Begrunnelse for at ekstern virksomhet har behov for å få overført eiendomsrett til resultatene:

6. Godtgjøring ved patenterbare oppfinnelser

Dersom studenten i forbindelse med utførelsen av oppgaven har nådd frem til en patenterbar oppfinnelse, enten alene eller sammen med andre, kan den eksterne kreve retten til oppfinnelsen overført til seg. Dette forutsetter at utnyttelsen av oppfinnelsen faller inn under den eksterne sitt virksomhetsområde. I så fall har studenten krav på rimelig godtgjøring. Godtgjøringen skal fastsettes i samsvar med arbeidstakeroppfinnelsesloven § 7. Fristbestemmelsene i § 7 gis tilsvarende anvendelse.

7. NTNU sine rettigheter

De innleverte filer av oppgaven med vedlegg, som er nødvendig for sensur og arkivering ved NTNU, tilhører NTNU. NTNU får en vederlagsfri bruksrett til resultatene av oppgaven, inkludert vedlegg til denne, og kan benytte dette til undervisnings- og forskningsformål med de eventuelle begrensninger som fremgår i punkt 8.

8. Utsatt offentliggjøring

Hovedregelen er at studentoppgaver skal være offentlige.

Sett kryss

x	Oppgaven skal være offentlig

I særlige tilfeller kan partene bli enige om at hele eller deler av oppgaven skal være undergitt utsatt offentliggjøring i maksimalt tre år. Hvis oppgaven unntas fra offentliggjøring, vil den kun være tilgjengelig for student, ekstern virksomhet og veileder i denne perioden. Sensurkomiteen vil ha tilgang til oppgaven i forbindelse med sensur. Student, veileder og sensorer har taushetsplikt om innhold som er unntatt offentliggjøring.

Oppgaven skal være underlagt utsatt offentliggjøring i (sett kryss hvis dette er aktuelt):

Sett kryss		Sett dato
	ett år	
	to år	
	tre år	

Behovet for utsatt offentliggjøring er begrunnet ut fra følgende:

Dersom partene, etter at oppgaven er ferdig, blir enig om at det ikke er behov for utsatt offentliggjøring, kan dette endres. I så fall skal dette avtales skriftlig.

Vedlegg til oppgaven kan unntas ut over tre år etter forespørsel fra ekstern virksomhet. NTNU (ved instituttet) og student skal godta dette hvis den eksterne har saklig grunn for å be om at et eller flere vedlegg unntas. Ekstern virksomhet må sende forespørsel før oppgaven leveres.

De delene av oppgaven som ikke er undergitt utsatt offentliggjøring, kan publiseres i NTNUs institusjonelle arkiv, jf. punkt 4, siste avsnitt. Selv om oppgaven er undergitt utsatt offentliggjøring, skal ekstern virksomhet legge til rette for at studenten kan benytte hele eller deler av oppgaven i forbindelse med jobbsøknader samt videreføring i et master- eller doktorgradsarbeid.

9. Generelt

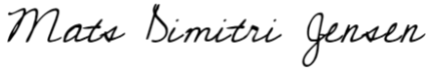
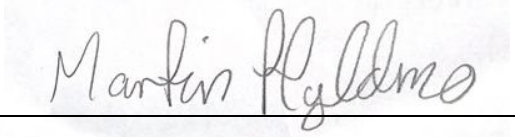
Denne avtalen skal ha gyldighet foran andre avtaler som er eller blir opprettet mellom to av partene som er nevnt ovenfor. Dersom student og ekstern virksomhet skal inngå avtale om konfidensialitet om det som studenten får kjennskap til i eller gjennom den eksterne virksomheten, kan NTNUs standardmal for konfidensialitetsavtale benyttes.

Den eksterne sin egen konfidensialitetsavtale, eventuell konfidensialitetsavtale den eksterne har inngått i samarbeidprosjekter, kan også brukes forutsatt at den ikke inneholder punkter i motstrid med denne avtalen (om rettigheter, offentliggjøring mm). Dersom det likevel viser seg at det er motstrid, skal NTNUs standardavtale om utføring av studentoppgave gå foran. Eventuell avtale om konfidensialitet skal vedlegges denne avtalen.

Eventuell uenighet som følge av denne avtalen skal søkes løst ved forhandlinger. Hvis dette ikke fører frem, er partene enige om at tvisten avgjøres ved voldgift i henhold til norsk lov. Tvisten avgjøres av sorenskriveren ved Sør-Trøndelag tingrett eller den han/hun oppnevner.

Denne avtale er signert i fire eksemplarer hvor partene skal ha hvert sitt eksemplar. Avtalen er gyldig når den er underskrevet av NTNU v/instituttleder.

Signaturer:

Instituttleder: Dato:	
Veileder ved NTNU: Dato: 19.01.23	
Ekstern virksomhet: Dato: 13.01.23	
Student: Dato: 13.01.23	
Student: Dato: 13.01.23	
Student: Dato: 13.01.23	
Student: Dato: 13.01.23	

Vedlegg B

Prosjektplan

NTNU
Kunnskap for en bedre verden

Institutt for informasjonssikkerhet og kommunikasjonsteknologi

DCSG2900 - Bacheloroppgave Bachelor i digital infrastruktur og cybersikkerhet

Prosjektplan

Forfattere:
Marte Jørgensen
Martin Hyldmo
Mats Dimitri Jensen
Thugitha Kanavathi

11.01.2023

Innhold

1 Mål og rammer

1.1 Bakgrunn

Denne bacheloroppgaven er det siste prosjektet vi som bachelorstudenter i digital infrastruktur og cybersikkerhet gjør. Oppdragsgiveren vår jobber i Digitaliseringsdirektoratet (Digdir), der han også gjennom dem og NTNU tar en off. PhD. Det er i forbindelse med hans forskning om avanserte statsfinansierte hackergrupper at denne oppgaven ble laget. Vi skal opparbeide kunnskap og en dypere innsikt om det mørke nettet TOR, i tillegg skal vi utvikle en crawler for å forsøke å høste større mengder med data fra det mørke nettet om ulike hackere og hackergrupper, der vi vil fokusere på å lete der det kan tenkes statsfinansierte hackergrupper befinner seg.

Det mørke nettet består av ganske store mengder med data noen ønsker å holde skjult. For å nå det mørke nettet trenger man en spesiell programvare og det er ikke noe man finner tilfeldig på "det vanlige" internett. Det er gjerne "The Onion Router" (TOR) som er mest kjent og som blir utgangspunktet for datainnhentingen i denne bacheloroppgaven. Det finnes flere grunner til å tro at mange av hackergruppene holder til på det mørke nettet og det er blant annet det vi ønsker å prøve å finne ut av. Det finnes enorme mengder data på det mørke nettet og vi skal prøve å få lastet ned mest mulig som omhandler hackere. All data vi laster ned vil overleveres til oppdragsgiver etter levert bacheloroppgave.

1.2 Prosjektmål

Vi har valgt å dele prosjektmål inn i tre underkategorier; effektmål, resultatmål og læringsmål. Effektmål handler om hva oppdragsgiver ønsker å oppnå med denne oppgaven [1]. Resultatmål handler om hva prosjektet skal levere til slutt og hva som er hovedpunktene [1]. Læringsmål går ut på det vi skal lære i løpet av dette prosjektet.

1.2.1 Effektmål

- Få oversikt over oppbyggingen, bakgrunnen og historien til TOR, samt data og datamengder som ligger der.
- Få hentet ut mest mulig informasjon om hackergrupper fra det mørke nettet via TOR, slik at det kan forskes videre på.
- Utvikle en crawler som høster dataene nevnt over.

1.2.2 Resultatmål

- Å ha en leveringsklar og oversiktlig database med ulike data fra det mørke nettet.
- Ha en ferdigutviklet crawler som kan hente ned innhold fra det mørke nettet relatert til nøkkelord man vil lete etter.
- Gjennomføre bacheloroppgaven på en sikker måte, slik at ingen på gruppen blir eksponert på det mørke nettet.

1.2.3 Læringsmål

- Opparbeide kunnskap om det mørke nettet, TOR, hvordan det fungerer og henger sammen.
- Opparbeide kunnskap om hvordan man kan holde seg sikker og anonym på det mørke nettet.
- Å gjennomføre større prosjektarbeid og metodisk arbeid.
- Hvordan lage en crawler og hvordan den fungerer.

1.3 Rammer

Rammene her er krav vi må forholde oss til i løpet av prosjektet, gitt av oppdragsgiver og NTNU. Disse skal ikke fravikes fra.

1.3.1 Tidsfrist

Denne prosjektplanen og standard kontrakt underskrevet av oss, veileder og oppdragsgiver skal leveres innen 31.01.23 på BlackBoard.
Bacheloroppgaven skal leveres senest 22.05.2023.
For studenter som skriver bachelor i digital infrastruktur og cybersikkerhet, som oss, skal vi skrive et individuelt refleksjonsnotat som skal leveres innen 30.05.23 også på Blackboard.

1.3.2 The Onion Router (TOR)

Det finnes flere måter å få tilgang til det mørket nettet, men i dette prosjektet er det kun TOR vi skal utforske. TOR bruker sikre og krypterte protokoller for å sikre brukerens anonymitet og var opprinnelig utviklet av og for det amerikanske militæret for å sikre sensitiv statskommunikasjon. TOR er nå tilgjengelig for alle og det er gratis [2]. TOR er også det mest kjente og det mest brukte når det gjelder det mørke nettet, noe som er blant årsakene til at det er TOR vi er bedt om å utforske og lære om.

1.3.3 Anonymitet

Alt som gjøres på det mørke nettet kan man forvente er overvåket, det er derfor imperativt at vi holder oss anonyme når vi gjennomfører eksperimentet. Det sies at det finnes tre typer brukere av det mørke nettet, det er de som har noe å skjule, etterforskere og sikkerhetsforskere (som vi klassifiseres som). Bruken av TOR vil anonymisere internettrafikken, dette gjøres ved å benytte seg av randomiserte noder med ulik geografisk lokasjon [3]. Startnoden vil fortsatt kunne knyttes geografisk til oss med mindre vi bruker VPN. For ekstra sikkerhet kan man kjøre VPN fra et annet system enn der man kobler seg opp til TOR [4].

2 Omfang

2.1 Problemområde

I følge "Threat Landscape" rapporten til ENISA fra 2022 så kan de mest utbredte angrepene spores tilbake til trusselaktører i disse fire kategoriene; statsfinansierte aktører, cyberkriminelle, hacktivister og hacker-for-hire aktører [5]. Statsfinansierte hackergrupper har blitt mer avanserte og ressursrike i løpet av de siste årene. Det er ikke lenger bare store bedrifter og nasjoner som er målgruppen, trender viser også at små og mellomstore bedrifter og enkeltpersoner er mer utsatt enn noen gang før [5]. Dette er noe som havner i medias søkelys stadig vekk, spesielt nå i forbindelse med Ukraina-krigen og den hybride krigføringen til Russland.

Mange statsfinansierte aktører har det mørke nettet som sitt naturlige habitat [6]. Det mørke nettet er et resultat av misbrukt anonymitet ved hjelp av en form for ruting, kalt onion routing. Onion routing har gjort det mulig å oppnå full anonymitet på kommunikasjonskanaler, og er vanskelig å spore tilbake til enkeltpersoner. I utgangspunktet ble onion routing utviklet for legitime brukere, slik at både avsender og mottagers identitet kunne beskyttes, som for eksempel militæret [7].

Hensikten med denne oppgaven, og dermed problemstillingen er å høste data som kan potensielt avdekke avanserte statlige trusselaktører, og gi innsikt i deres form for kommunikasjon og operasjonelle plan. Vi må sette oss inn i hvordan TOR fungerer og er bygd opp for å klare å navigere oss rundt og finne chatrom, søkemotorer og protokoller som brukes. Vi skal utvikle en crawler som laster ned dataene vi leter etter, og eventuelle funn skal lagres i en egen database og skal kunne brukes til forskning i etterkant.

2.2 Problemavgrensning

Det finnes flere mørke nett, men oppgaven er avgrenset til å kun se på det mest kjente "The Onion Router". Det vil si at oppgaven baseres kun på data som er tilgjengelig gjennom nettleseren TOR og nettsider med .onion adresser. Hovedfokuset vil være å høste mest mulig data som kan knyttes til hackergrupper, ideelt sett vil noe av informasjonen omhandle statsfinansierte trusselaktører, som oppdragsgiver kan bruke i videre forskning. Oppgaven innebærer ikke at vi skal gå gjennom alt og kategorisere trusselaktøren basert på informasjonen, dette forblir oppdragsgivers ansvar. I forkant av dette vil vi sette oss inn i hvordan TOR fungerer og er satt sammen, slik at det blir enklere å navigere seg rundt med tanke på å finne ønsket data.

Vi skal kjøre crawleren i nesten to måneder og vil dermed kun overlevere data som vi har klart å høste på denne tiden. Det vil si at alle nettsider som dukker opp etter denne tidsperioden ikke vil bli inkludert. Nettsidene som er tatt ned eller sperret av diverse årsaker vil vi heller ikke ha mulighet til å hente informasjon fra.

2.3 Problemstilling

Grunnet omfanget av oppgaven har vi valgt å ikke formulere problemstillingen som et forskningsspørsmål. Som nevnt under problemområde, vil hovedmålet være å få en oversikt over hvilke typer data, og datamengder som ligger på TOR nettet. Informasjon om avanserte statlige trusselaktører vil være

det mest relevante for vår oppdragsgiver. For å finne disse aktørene vil det være vesentlig å tilegne seg kunnskap om hvordan TOR er bygd opp og navigasjon inne på det mørke nettet først. Fokuset vårt vil likevel være å høste data uavhengig av trusselaktør, og analyseringen overlates til oppdragsgiver. Informasjonen som høstes skal lagres i en egen database som overleveres til oppdragsgiver i mai.

3 Prosjektorganisering

3.1 Ansvarsforhold

Leder: Lederen er ansvarlig for prosjektets fremgang. Hvis prosjektarbeidet skulle stoppe opp, er det lederen som har ansvaret for å få igang arbeidet igjen.

Sekretær: Sekretæren er ansvarlig for å notere ned møtereferat etter møter med oppdragsgiver og veileder. Møtereferatet legges inn i et fellesdokument for alle møtereferatene.

Dokumentansvarlig: Dokumentansvarlig er ansvarlig for å kvalitetssikre dokumenter, samt kvalitetssikre kode og diagrammer. Han skal også sikre for at alt arbeid som er ferdigstilt er av ønsket kvalitet.

Ansvarlig rombooking: Ansvar for å booke rom både til veiledning og møte med arbeidsgiver. I tillegg til gruppearbeid.

Kildeansvarlig: Sørge for at kildene er riktige og at bibliografi er i orden.

Alle: Enhver er ansvarlig for å registrere egne timer i et felles Excel-ark. Når man har brukt en kilde, legges kildens detaljer inn i kilde-filen i Overleaf.

3.2 Roller

- Leder - Mats
- Sekretær - Alle (rullering)
- Dokumentasjonsansvarlig - Martin
- Ansvarlig rombooking - Thugitha og Mats
- Kildeansvarlig - Marte

3.3 Rutiner

- Timeføring skal være ført inn i timelisten i Teams, dette skal være oppdatert hver uke, i tillegg en kort kommentar om hva som er jobbet med.
- Planlagte statusmøter med oppdragsgiver er i hovedsak 13:00 - 13:30 på fredager.
- Planlagte møter med veileder er i hovedsak 10:15 - 11:00 på torsdager.
- Gruppen samles til felles arbeidsøkter mandager, onsdager/torsdag og fredager. Avvik fra dette kan forekomme da det kan være forelesninger eller andre møter i denne tiden.
- Vi bruker GitLab til å holde orden på kode, samt planlegging av koden. Her kan gruppen hente ned laget kode, og laste opp ny kode man har laget.

3.4 Grupperegler

- Oppmøte for felles arbeidsøkter er enten 08.15 eller 9:15.
- Ved forsinkelser skal andre gruppemedlemmer informeres.
- Oppgaver fordeles likt, slik at alle bidrar like mye.
- Oppgaver skal ferdigstilles innen bestemt tid, sett ut ifra Gantt-diagrammet. Hvis dette ikke er mulig, må det informeres om i god tid før bestemt ferdigstillingsdato.
- Når vi jobber, skal det være fullt fokus på prosjektarbeidet.
- Fast lunsjpause på felles økter, fra 12.00 - 12.30. Små pauser innimellom tas på sparket.

4 Planlegging, oppfølging og rapportering

4.1 Valg av prosessrammeverk

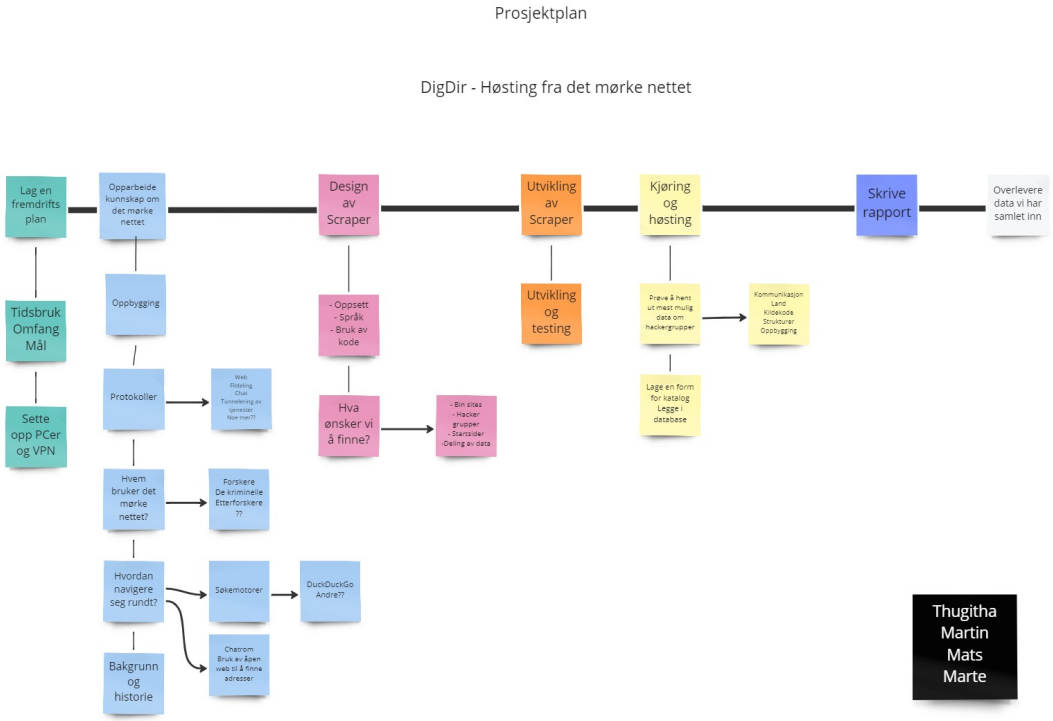
Valg av prosessrammeverk var ikke enkelt. Vår bacheloroppgave er en blanding mellom en god del research og i tillegg utvikling av en crawler, samt å lage en database der nedlastet data skal legges inn. Både Scrum og Kanban var alternativer som ble presentert for oss, men vi følte ikke at noen av de passet helt for vår type oppgave. Derfor fant vi ut at metoden Scrumban som kombinerer de to prosessrammeverkene kanskje var en bedre tilnærming i vårt tilfelle. Fordelen med Scrumban er at man kan spare tid ved å få en oversikt over prosjektstyringen, man ser hva som skal gjøres og allerede er gjort. Scrumban er også fint for litt langvarige prosjekter, da det er mulig å gjøre endringer underveis om nødvendig. Scrumban har ingen spesifikke roller slik som Scrum har, dette gjør teamet mer autonomt. Her kan man styre endring av prioritet, slik at man gjør det som trengs først uten at man bare må følge en statisk plan. Her er det medlemmene selv som velger hva de vil gjøre uten å få tildelt en oppgave av en Scrum master for eksempel [8].

Selv om et prosessrammeverk uten en ledelse kan virke forvirrende er vi et såpass lite og tett team at vi ser på det som mindre tidkrevende i dette prosjektet, uten at det vil gjøre det mer komplisert. Denne blandingen av Scrum og Kanban er relativt ny, men vi finner likevel nok informasjon for å tilpasse den til vårt prosjekt. Produkteier har mindre kontroll, men det har ikke noe å si for vår del, da vi har faste møter med oppdragsgiver hver uke, der vi deler hva vi har gjort og skal gjøre. Det går derfor fint at vi har et prosessrammeverk som hovedsaklig er tilpasset gruppen og de oppgavene vi skal utføre [8]. I tillegg er det fleksibel lengde på iterasjonene (det som skal gjøres), som passer fint da ulike oppgaver vil ta forskjellig lengde med tid.

4.1.1 Hvordan vi skal bruke Scrumban:

Vi bruker Trello for vårt Scrumban board, der vi har en backlog, en to-do/prioritering (det som må gjøres først), en "doing" liste som viser hvem som jobber med hva, som er greit når vi begynner å fordele oppgaver litt mer. Til slutt en "done", der man kan se hva som er fullført. Her vil man for eksempel kunne sette seg selv på en oppgave, slik at de andre kan se hvem som jobber med hva. Det er mulig å være flere om en oppgave, for eksempel at vi alle skriver denne prosjektplanen sammen. Siden vi uansett møtes minst 3 ganger i uken, trenger vi ikke noe eget Scrumban-møte, vi vil uansett oppdatere hverandre hver gang vi møtes.

Bildet under er en oversikt for å visualisere den prosessen vi har vært igjennom og for å vise hvordan vi har tenkt under planleggingen. Her kan man se hvordan vi tenker å bygge opp prosjektet, hvor vi ønsker å starte og rekkefølgen vi vil gjøre det i. Som nevnt skal vi bruke Scrumban og denne oversikten er det vi vil ta utgangspunkt i når vi skal lage Scrumban-boardet vårt.



4.2 Plan for statusmøter og beslutningspunkter i perioden

4.2.1 Statusmøter

Vi har valgt å ha ukentlige møter med oppdragsgiver. Disse vil holdes hver fredag kl. 13.00, og skal fungere som statusmøter, der vi oppdaterer oppdragsgiver og reflekterer over fremdriften. Dette gir oss også muligheten til å stille oppdragsgiveren spørsmål som kan ha dukket opp i løpet av uka og får råd og veiledning dersom vi skulle stå fast. Faste møter med veileder er hver torsdag kl. 10.15. Dette gir oss mulighet til å reflektere med veileder før møtet med oppdragsgiver, samt å diskutere relevante temaer rundt prosjektet. Vi har også kommet til enighet om å ha faste møter hver mandag kl.09.15 internt, slik at vi kan få en oversikt over hva som må gjøres den kommende uka og fordele arbeidsoppgaver.

4.2.2 Beslutningspunkter

Vi har tre faste dager i uka der vi møtes på campus og jobber som en gruppe, derfor vil det være rom for å diskutere ulike synspunkter før eventuelle beslutninger tas. Dersom det skulle dukke opp viktige avgjørelser i de dagene vi ikke jobber sammen så har vi laget en Teams gruppe til dette formålet slik at det er enkelt å komme i kontakt med hverandre, i tillegg til at vi har andre kommunikasjonsplattformer som Messenger og Discord. Det vil også være en selvfølge å ta beslutninger fortløpende gjennom hele prosjektet og ikke minst når vi skal inn i nye faser.

5 Organisering av kvalitetssikring

5.1 Dokumentasjon og verktøy

Dokumenter som gruppen lager lagres i et Teams team som kun medlemmene i gruppen har tilgang til. Dette gjelder da timelister, møtereferat, generelle notater, samt prosjektplanen og bachelorrapporten. Dette teamet i Teams er beskyttet av multifaktor-autentisering gjennom Feide, slik at dette teamet er et sikkert sted å lagre dokumenter. Etterhvert som prosjektplanen og bachelorrapporten blir ferdigstilte vil disse lagres på en minnepinne.

Både prosjektplanen og bachelorrapporten skrives i LaTex gjennom redigeringsprogrammet Overleaf. Her er det kun gruppemedlemmene som har tilgang. På denne måten vil både bachelorrapporten og prosjektplanen være lagret på tre steder, Overleaf, Teams og en minnepinne.

For å holde orden på utviklingen og testing av programmet vi skal lage, bruker vi GitLab-området til NTNU IDI. Her kan alle gruppemedlemmer lett hente ned nylige oppdateringer og lett gjøre oppdateringer selv. På GitLab kommer vi også til å benytte 'Issues' funksjonen til å fordele oppgaver tilknyttet utvikling og til innhenting av informasjon rundt crawleren. Gruppen kommer til å benytte seg av Visual Studio Code (VSC) som IDE til å skrive og teste koden som skal danne crawleren. Ved bruk av VSC kan gruppemedlemmene enkelt laste ned utvidelsespakker som gjør selve kodingen lettere, samtidig som VSC er en IDE som støtter de aller fleste programmeringsspråk.

For å holde oss helt anonyme på det mørke nettet vil koble oss til en VPN før vi går på TOR nettleseren. Vi vil bruke en raspberry pi som ruter, der denne alltid vil være koblet på Ekspress-VPN.

Oppdragsgiver har utstyrt gruppen med fire PCer, som skal brukes til høstingsaktiviteten fra det mørke nettet. Disse PCene er tilbakestilt slik at det ikke er noen spor som kan brukes til identifisering av noen av gruppens medlemmer, eller tidligere eiere av PCene. Oppdragsgiver vil skaffe harddisker som etter planen skal brukes til å ha nedlastet data fra det mørke nettet på.

MySQL vil bli brukt som database for å få et ryddig system på linkene og innhold som vi vil ta vare på. Vi vil ha en maskin som kjører MySQL, med en API som gjør at maskinene som scraper det mørke nettet vil kunne sende data over.

5.2 Plan for inspeksjoner og testing

Under vårt prosjekt inngår det å utvikle og kjøre en crawler, denne trenger ingen inspeksjon fra oppdragsgiver/veileder. Testing av crawleren vil foregå fortløpende under utviklingen, og kun skje internt i gruppa. Vi vil selv revidere koden ut fra utfallene som hver test gir.

5.3 Risikoanalyse

Gruppen har valgt å gjøre en risikoanalyse med ulike scenarioer som kan oppstå i i løpet av prosjektarbeidet. For hvert av risikoscenarioene har vi utarbeidet tiltak slik at vi har klare løsninger dersom hendelsen skulle oppstå, i tillegg har vi vært proaktiv og lagt føringer for å minimere risikoen

for flere av disse scenarioene. Tabellen nedenfor viser mål på både sannsynlighet og konsekvens, hvor hver av parameterene er delt inn i fem nivåer. Disse fem nivåene bygger på den samlede risikoen som er definert med de fargede feltene i tabellen, hvor rød er kritisk, oransje er høy, gul er middels og grønn er lav. Med disse målene for samlet risiko kan vi beregne hvilke scenarioer som er viktigst å løse. Scenarioene med høyest samlet risiko har flest tiltak.

5.3.1 Risikomatrise

	Konsekvens				
Sannsynlighet	Svært høy	Stor	Medium	Liten	Ingen
Svært sannsynlig					
Meget sannsynlig				12	
Sannsynlig		1, 10			
Mindre sannsynlig	4, 6	9		11, 3	
Usannsynlig	2, 7	5		8	

5.3.2 Risikoscenarioer

Risiko — Beskrivelse	
1,	Laster ned virus/malware som kan ødelegge data allerede lastet ned
2,	Datatap i form av rapport/dokumenter/vedlegg
3,	Får ikke ut større mengde data fra det mørke nettet
4,	Langvarig sykdom for ett/flere medlemmer eller hos nær familie
5,	Et medlem forlater gruppa
6,	Anonymiteten kompromitteres
7,	Ikke ferdig til tidsfristen
8,	En/flere følger ikke grupperegler/konflikt
9,	Ikke nok teori/informasjon å finne på det åpne nettet
10,	Går utenfor scope
11,	Faglig innhold ikke som forventet
12,	Utfordrende å tolke oppbyggingen av nettverket (TOR)

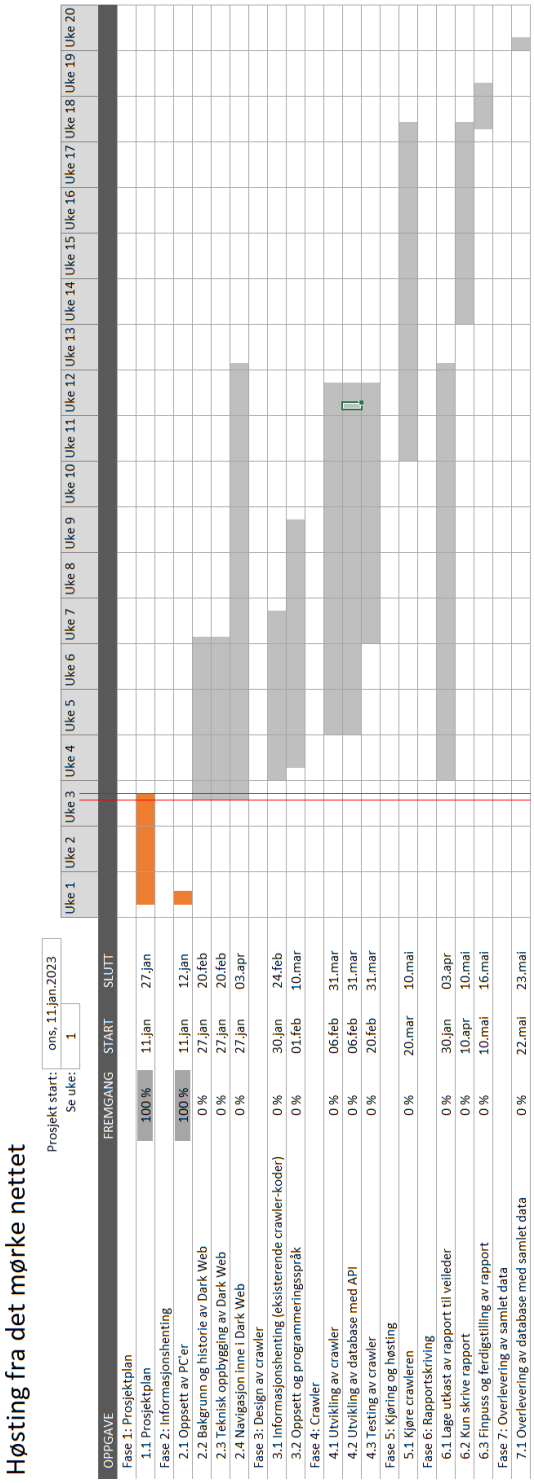
5.3.3 Tiltak for risikoscenarioer

Samlet risiko	Nr.	Tiltak
Høy	1	Som en forutsetning vil vi legge inn en automatisk jevnlig overføring av nedlastet data til en ekstern harddisk. Dersom det likevel skulle skje må vi tilbakestille den aktuelle PCen som har blitt infisert.
Middels	2	Viktige dokumenter som dette blir lagret på Overleaf, samt lastes ned lokalt av hvert gruppemedlem etter hver økt.
Lav	3	Starter utvikling, testing og kjøring med crawler tidlig for å øke sannsynligheten for å finne mest mulig data. Konsekvensen er liten, da det er mye variert data som er relevant. Oppdragsgiver har sagt at til og med å ikke finne er et funn.
Høy	4	Lite sannsynlig, men dersom det skulle skje tar vi hensyn til det om det skulle oppstå. Alt kan gjøres digitalt så det er en mulighet for tilpasning.
Middels	5	Siden vi er 4 stykker, vil det ikke ha alt for stor konsekvens ved at en forlater, vi er alle innstilt på å stå løpet ut. Dersom det skulle skje er alle klar for å ta i et ekstra tak.
Høy	6	Vi skal bruke PCer vi har fått som er helt rensket og i tillegg bruke VPN. TOR bruker en protokoll som gjør det vanskelig å spore data tilbake til bruker.
Middels	7	Vi jobber etter satt tidsskjema, startet tidlig og jobber jevnlig. Vi alle er villige til å jobbe både i helger og ferier for å komme i mål.
Lav	8	Tas opp internt om det skulle forekomme, vi ønsker at det skal være enkelt å si ifra om alt. Prosjektleder tar dette opp med den enkelte dersom man ikke klarer å løse det på gruppenivå.
Middels	9	Finne andre kilder som bøker, intervjuobjekter eller eventuelt samle så mye informasjon som mulig på det mørke nettet.
Høy	10	Viktig at vi har problemavgrensningen i bakhodet. I tillegg jevnlig kontakt med oppdragsgiver og veileder slik at eventuelle avvik kan fanges opp.
Lav	11	For å unngå dette har vi brukt tid i starten til å planlegge prosjektet og avgrensningen for å få ønsket innhold. Dersom dette skulle oppstå må vi ta et ekstra møte med oppdragsgiver og veileder.
Middels	12	Det kan for eksempel være flere lag med sikkerhet, det kan være forsøk på å gjemme seg eller det kan være språkvansker (i form av ulike språk som brukes på det mørke nettet)

6 Plan for gjennomføring

6.1 Gantt-skjema

Gantt-skjemaet vi har laget er et interaktivt skjema, hvor gruppens medlemmer endrer på kolonnen markert med 'Fremgang'. Dette er for å ha en oversikt over fremgangen til de ulike oppgavene. Vi har satt hver hoveddel som sine egne faser, slik at hver fase også kan fungere som en milepæl.



OPPGAVE	FREMGANG	START	SLUTT
Fase 1: Prosjektplan			
1.1 Prosjektplan	100 %	11.jan	27.jan
Fase 2: Informasjonshenting			
2.1 Oppsett av PC'er	100 %	11.jan	12.jan
2.2 Bakgrunn og historie av Dark Web	0 %	27.jan	20.feb
2.3 Teknisk oppbygging av Dark Web	0 %	27.jan	20.feb
2.4 Navigasjon inne i Dark Web	0 %	27.jan	03.apr
Fase 3: Design av crawler			
3.1 Informasjonshenting (eksisterende crawler-koder)	0 %	30.jan	24.feb
3.2 Oppsett og programmeringsspråk	0 %	01.feb	10.mar
Fase 4: Crawler			
4.1 Utvikling av crawler	0 %	06.feb	31.mar
4.2 Utvikling av database med API	0 %	06.feb	31.mar
4.3 Testing av crawler	0 %	20.feb	31.mar
Fase 5: Kjøring og høsting			
5.1 Kjøre crawleren	0 %	20.mar	10.mai
Fase 6: Rapportskriving			
6.1 Lage utkast av rapport til veileder	0 %	30.jan	03.apr
6.2 Kun skrive rapport	0 %	10.apr	10.mai
6.3 Finpuss og ferdigstilling av rapport	0 %	10.mai	16.mai
Fase 7: Overlevering av samlet data			
7.1 Overlevering av database med samlet data	0 %	22.mai	23.mai

Bibliografi

[1] *Mål*. SNL. 24th Nov. 2022. URL: https://snl.no/m%C3%A5l_-_prosjektledelse (visited on 16/01/2023).

[2] Jake Frankenfield. *What Is Tor? Who Uses It, How to Use It, Legality, and Purpose*. Investopedia. 23rd Sept. 2022. URL: https://www.investopedia.com/terms/t/tor.asp (visited on 26/01/2023).

[3] *Løk-ruting*. SNL. 5th Aug. 2022. URL: https://snl.no/l%C3%B8k-ruting (visited on 30/01/2023).

[4] *TorPlusVPN*. TORProject. 15th June 2020. URL: https://gitlab.torproject.org/legacy/trac/-/wikis/doc/TorPlusVPN (visited on 30/01/2023).

[5] 'ENISA Threat Landcape 2022'. In: (3rd Nov. 2022), pp. 22–23. URL: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2022 (visited on 16/01/2023).

[6] CrowdStrike. 'EXPOSING THE OPEN, DEEP, AND DARK WEB AND BEYOND'. In: (2021), p. 3.

[7] *Det mørke nettet*. SNL. 1st Aug. 2022. URL: https://snl.no/Det_m%C3%B8rke_nettet (visited on 16/01/2023).

[8] Sarah Laoyan. *Scrumban: The best of two Agile methodologies*. 6th June 2022. URL: https://asana.com/resources/scrumban (visited on 20/01/2023).

Vedlegg C

Oppgavebeskrivelse

Oppgavebeskrivelse levert av Digdir.

Høsting fra «det mørke nettet»

Jeg tar en Off. PHD gjennom Digitaliseringsdirektoratet, som har tjenester for dialog mellom organisasjoner, virksomheter og individ med det offentlige. De mest kjente tjenestene som Digitaliseringsdirektoratet tilbyr, er Altinn og id-porten. I tillegg til 20 andre offentlig IKT løsninger.

Oppgaven

Det mørke nettet består av store mengder med data, som noen ønsker å skjule. Forskjellen på «Dark Web», og «Deep Web», så er Deep Web alt som ligger bak en autentiseringsløsning på Internet, men Dark Web krever spesialprogramvare for å få tilgang til data som ligger det. Det finnes flere mørke nett. Men det mest kjente er «The Onion Router» , TOR , som har adresser som slutter på .onion, og krever TOR nettleser for å få tilgang. Ønsket er å få hentet ut data fra det mørke nettet til å forske videre på.

Oppgavens mål

Oppgaven sitt mål er å få en oversikt over hvilke typer data, og datamengder som ligger på TOR nettet. Oppgaven er teknisk i så måte at dere må sette dere inn i hvordan TOR fungerer, og ikke minst finne , eventuelt utvikle en crawler som kan hente ned data for dere. **Det er også imperativt viktig at dere holder dere anonyme når dere gjennomfører eksperimentet.** Alt skal selvsagt dokumenteres , men alt som gjøres på det mørke nettet må forventes å være overvåket. Det sies at det finnes bare tre typer brukere på det mørke nettet. De som har noe å skjule, de som etterforsker, og sikkerhetsforskere. Dere vil være sikkerhetsforskere.

Oppgavens krav

- Forstå hvordan det mørke nettet er bygget opp
- Forstå hvilke forhåndsregler som kreves for å være anonym
- Utvikle en måte å høste data fra ulike sider på det mørke nettet.
- Laste innholdet som høstes over i en database for «offline» bruk.

Siden jeg arbeider med en PHD om avanserte statsfinansierte hackergrupper, så kjenner jeg til at en del av slike gruppers «naturlige habitat» er på det mørke nettet. Ønsker derfor å ha mulighet til å kunne bruke eventuelle funn til å gjøre analyser på i forhold til forskningen min.

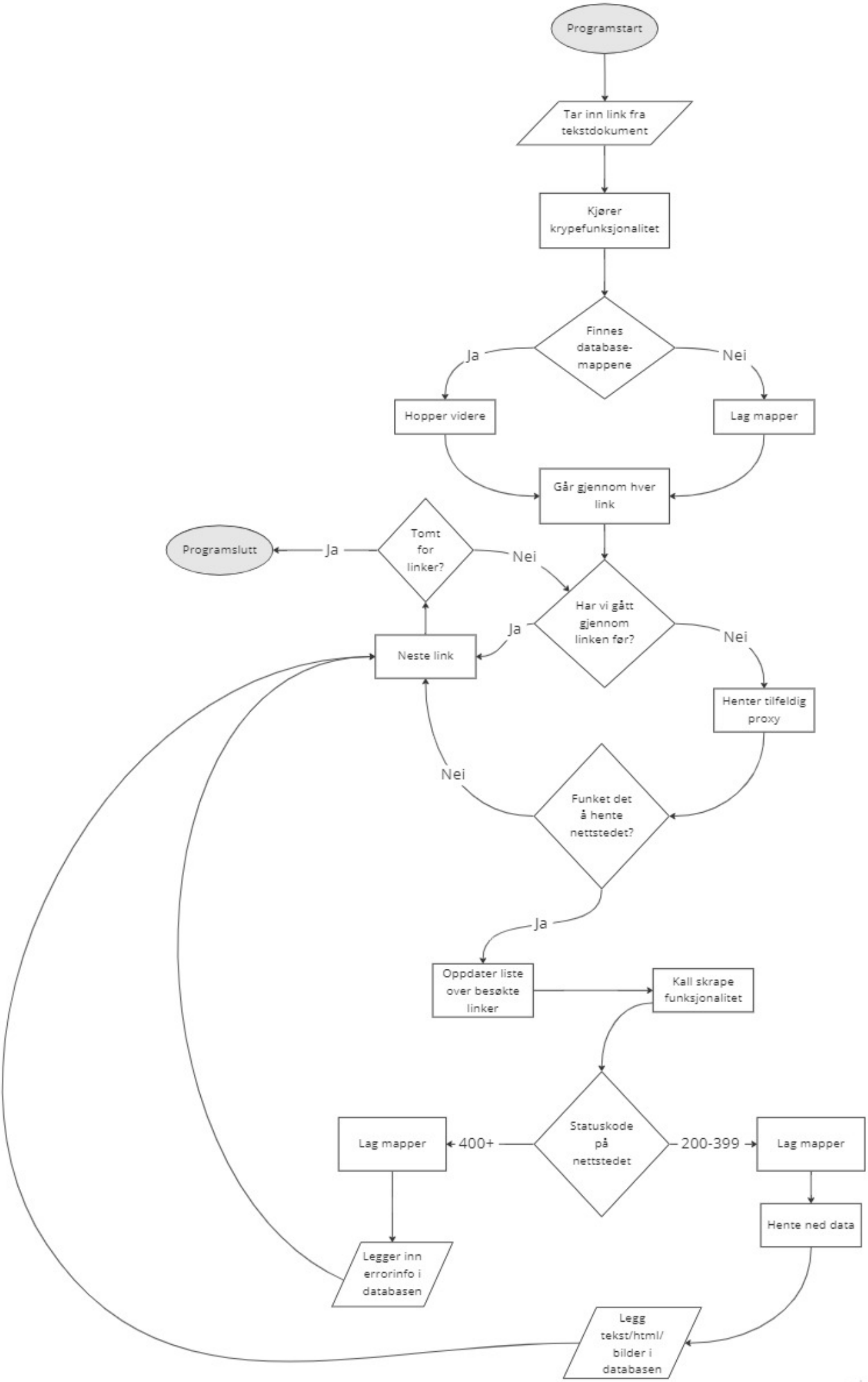
Raymond Andre Hagen

raymohag@stud.ntnu.no / raymond.andre.hagen@digdir.no

Vedlegg D

Flytdiagram av versjon 1 av søkeroboten

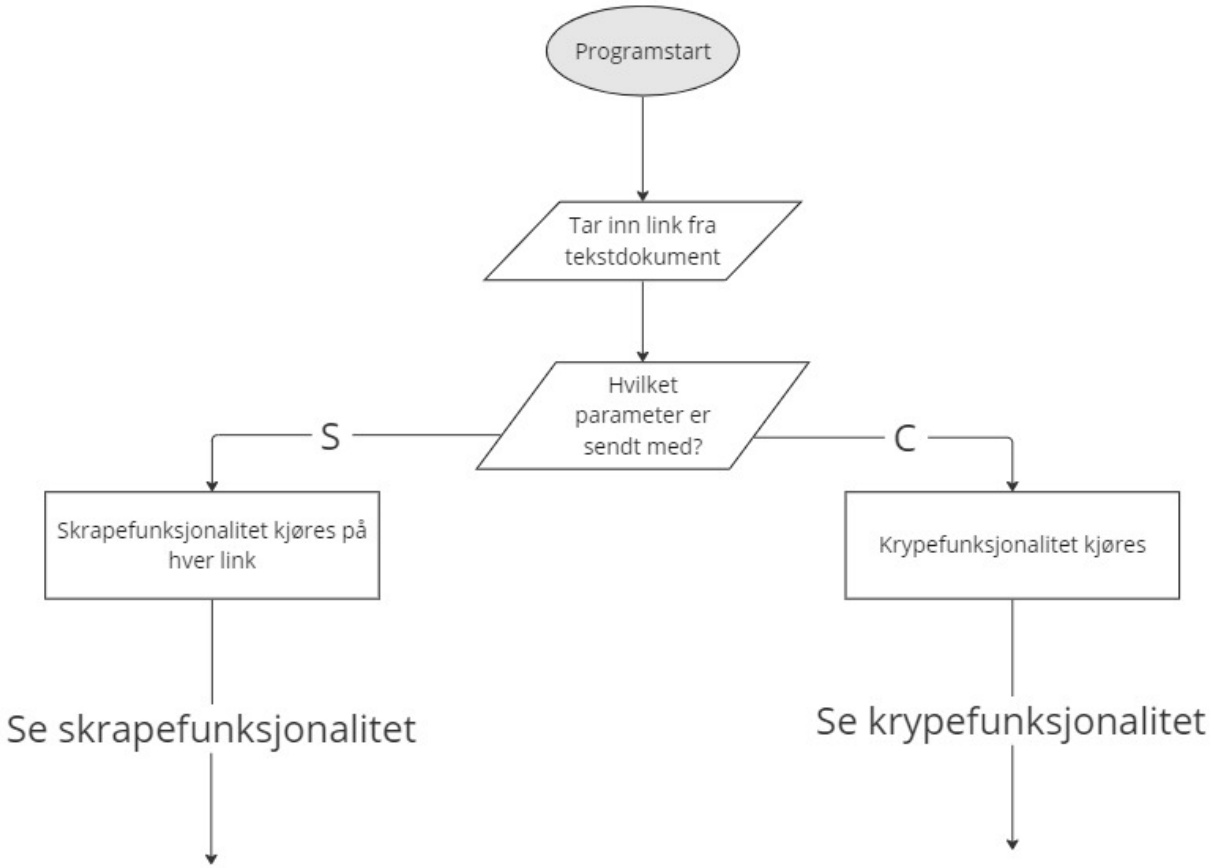
Flytdiagrammet viser programflyten for versjon 1 av søkeroboten.

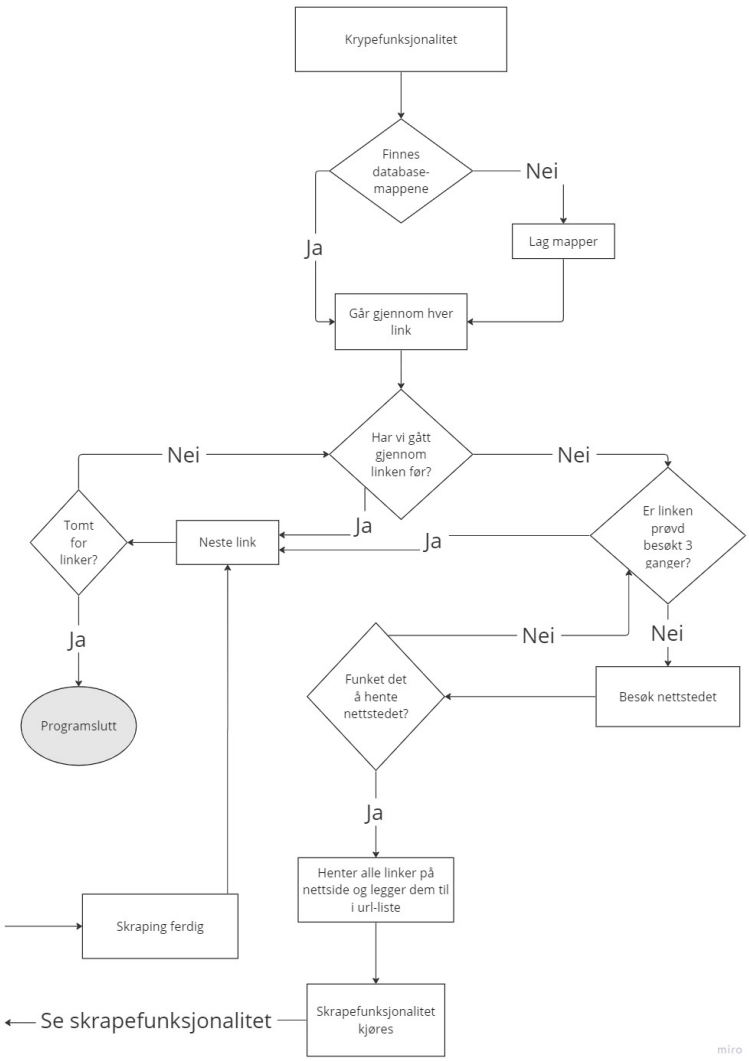

Programstart
Tar inn link fra tekstdokument
Kjører krypefunksjonalitet
Finnes database-mappene
Ja
Nei
Hopper videre
Lag mapper
Går gjennom hver link
Tomt for linker?
Ja
Programslutt
Nei
Har vi gått gjennom linken før?
Ja
Nei
Neste link
Henter tilfeldig proxy
Nei
Funket det å hente nettstedet?
Ja
Oppdater liste over besøkte linker
Kall skrape funksjonalitet
Statuskode på nettstedet
400+
200-399
Lag mapper
Lag mapper
Hente ned data
Legger inn errorinfo i databasen
Legg tekst/html/ bilder i databasen

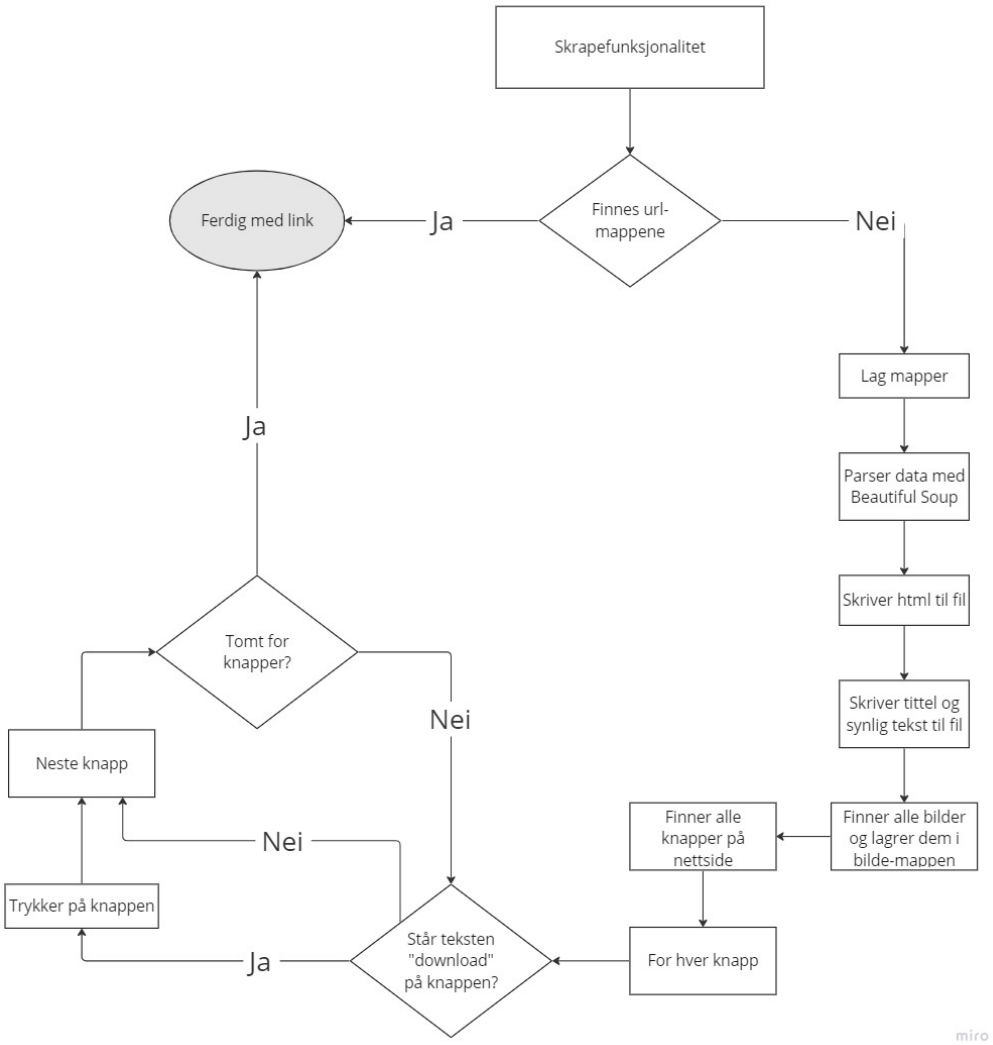
Vedlegg E

Flytdiagram av versjon 2 av søkeroboten

Flytdiagrammene viser programflyten for versjon 2 av søkeroboten.


Programstart
Tar inn link fra tekstdokument
Hvilket parameter er sendt med?
S
C
Skrapefunksjonalitet kjøres på hver link
Krypefunksjonalitet kjøres
Se skrapefunksjonalitet
Se krypefunksjonalitet


Krypefunksjonalitet
Finnes database-mappene
Nei
Lag mapper
Ja
Går gjennom hver link
Har vi gått gjennom linken før?
Nei
Nei
Tomt for linker?
Neste link
Ja
Ja
Er linken prøvd besøkt 3 ganger?
Ja
Nei
Nei
Programslutt
Funket det å hente nettstedet?
Besøk nettstedet
Ja
Henter alle linker på nettside og legger dem til i url-liste
Skraping ferdig
Skrapefunksjonalitet kjøres
Se skrapefunksjonalitet
miro


Skrapefunksjonalitet
Finnes url-mappene
Ja
Ferdig med link
Nei
Lag mapper
Parser data med Beautiful Soup
Skriver html til fil
Skriver tittel og synlig tekst til fil
Finner alle bilder og lagrer dem i bilde-mappen
Finner alle knapper på nettside
For hver knapp
Står teksten "download" på knappen?
Ja
Trykker på knappen
Nei
Neste knapp
Tomt for knapper?
Nei
Ja
miro

Vedlegg F

Søkeroboten

Kildekoden til den egenutviklede søkeroboten.

```
import sys
import getopt
import os
from selenium import webdriver
from selenium.webdriver.firefox.options import Options
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.firefox.service import Service
from webdriver_manager.firefox import GeckoDriverManager
from selenium.webdriver.firefox.service import Service
from selenium.webdriver.common.by import By
from selenium.common.exceptions import TimeoutException
from bs4 import BeautifulSoup
from bs4.element import Comment
import shutil
import time
from random import randint

urls = []
special_chars = ['#','<', '>', '+', '%', '!', '‘', '  ', '&', '*', '\'', '\"',
                 '|', '{', '}', '?', '=', '/', '\\', ':', '␣', '@']
crawledUrls = []

def main(argv, driver):
    inputfile = ''
    try:
        opts, args = getopt.getopt(argv, "hi:cs")
    except getopt.GetoptError:
        print('test.py␣-i␣<inputfile>')
        sys.exit(2)


    for opt, arg in opts:
        if opt == '-h':
            help = r"""
␣␣␣␣scraper.py␣-<>

␣␣␣␣-i␣<inputfile>␣"File␣containing␣urls␣to␣crawl/scrape␣from"
␣␣␣␣-s␣"Adds␣only␣scraping␣functionality␣to␣the␣program"
␣␣␣␣-c␣"Adds␣crawling␣and␣scraping␣functionality␣to␣the␣program"

␣␣␣␣␣␣␣␣␣␣␣␣"""
            print(help)
            sys.exit(2)
        elif opt == '-s':
            for url in urls:
                driver.get(url)
                scraper(driver)
        elif opt == '-c':
            crawler(driver)
        elif opt == '-i':
            inputfile = arg
            f = open(inputfile, 'r')
            url = f.readlines()
            for x in url:
                urls.append(x.strip('\n'))
    print('Input␣file␣is␣', inputfile)
```

```
def tag_visible(element):
    if element.parent.name in ['style', 'script', 'head', 'title',
                               'meta', '[document]']:
        return False
    if isinstance(element, Comment):
        return False
    return True



def scraper(driver):

    url = driver.current_url        # Getting the current URL
    uri = makeUri(url)              # Removing 'https://' and invalid characters

    # Checking if the URL is scraped before
    if os.path.exists(f"database/scraped/{uri}"):
        return      # Returning without making changes

    # Creating the directories for the current URL
    os.makedirs(f"database/scraped/{uri}/images", exist_ok=True)
    os.makedirs(f"database/scraped/{uri}/data", exist_ok=True)

    # Opening the files that will be written to,
    # both the available data and the raw html
    data = open(f"database/scraped/{uri}/data/data.txt", "a", encoding="utf-8")
    site = open(f"database/scraped/{uri}/website.html", 'w', encoding="utf-8")

    # Parsing the html with BS4
    soup = BeautifulSoup(driver.page_source, 'html.parser')

    # Write the html to a file with correct indentation
    site.write(BeautifulSoup(driver.page_source, 'html.parser').prettify())

    # Check for a title and write to the data file
    if soup.has_attr('title'):
        title = soup.find('title').get_text()

        if title:
                data.write(title)
                data.write("\n")

    # Find all visible text on the website and write this to the data file
    texts = soup.find_all(string=True)
    visible_texts = filter(tag_visible, texts)
    data.write(u"\n".join(t.strip() for t in visible_texts))

    # Wait for pictures to load, and get their html location
    print('Sleeping 12 - soon images')
    time.sleep(12)
    images = driver.find_elements(By.TAG_NAME, 'img')

    counter = 1

    # For all images try to screenshot and save image in the website's folder
    if images:
        for image in images:
            # Remove the svg extention images
```

```
            if (image and ('.svg' not in image.get_attribute('src'))):
                try:
                    while os.path.isfile(f'database/scraped/{uri}/images/image'
                                         + str(counter) + '.png'):
                        counter += 1

                    image.screenshot(f'database/scraped/{uri}/images/image'
                                     + str(counter) + '.png')
                    print('Sleeping 2 - between screenshots')
                    time.sleep(2)

                except Exception as e:
                    print(f'There is no such element: {e}!!')
                    continue

    # Get all the buttons on the website
    buttons = driver.find_elements(By.TAG_NAME, 'button')


    if buttons:
        # for every button on the website - if it starts a download - click it
        for button in buttons:
            print(button.text.upper())
            if 'DOWNLOAD' in button.text.upper():
                try:
                    button.click()
                    print("Sleep 200 - Downloading file")
                    time.sleep(200)

                except Exception as e:
                    print(f'Button is not working as expected {e}')

    moveFiles(f'database/scraped/{uri}/downloads/')




def crawler(driver):
    driver = driver
    crawlFile = open("finished.txt", "a")

    # Make all the folders to be used later
    if not os.path.exists("database"):
        os.makedirs("database", exist_ok=True)

        if not os.path.exists("database/scraped"):
            os.makedirs("database/scraped", exist_ok=True)


    for url in urls:
        retries = 0
        check = False

        # Check if the url has been visited before
        if url not in crawledUrls:

            # Max 3 retries for any fault from a website
            while(retries < 3 and not check):
                print("top of while")
                try:
```

```
                    print('trying to connect...')
                    driver.get(url)
                    print('Visiting: ' + url)
                    print('Sleeping 17 - soon scraping')
                    time.sleep(17)
                    check = True


                except TimeoutException as e:
                    print(f'Ignoring address lookup on {e}!!')
                    driver.close()
                    driver = startSelenium()
                    retries += 1

                except Exception as e:
                    print(f'Ignoring address lookup on {e}!!')
                    driver.close()
                    driver = startSelenium()
                    retries += 1

            # If it exceeded the limit of 3 retries - go to next link
            if retries == 3:
                continue

            # Append the visited url to the list of visited urls
            crawledUrls.append(url)

            # Send the driver to the scraper
            scraper(driver)

            # The same site will not be visited
            # again the next time the program starts
            crawlFile.write(url + '\n')

            links = driver.find_elements(By.TAG_NAME, 'a')


            # For all the links on the site -
            # check if it is in the list or have been visited -
            #  add it to the list of urls to be visited.
            for link in links:
                if link not in crawledUrls or link not in urls:
                    if link.get_attribute('href'):
                        getLink = link.get_attribute('href')

                        if 'http://' in getLink or 'https://' in getLink:
                            if getLink not in crawledUrls or getLink not in urls:
                                urls.append(getLink)
                                appendUrls()

                        elif getLink[0] == '/':
                            print(url+ getLink)
                            urls.append(url+ getLink)

        # Just a line separator
        print (f'-----------------{url}-------------------')
    print(crawledUrls)


def makeUri(uri):
```

```
    if 'https://' in uri:
        uri = uri[8:]
    else:
        uri = uri[7:]
    stringOk = uri
    for i in special_chars:
        stringOk = stringOk.replace(i, '_')

    return stringOk


def appendUrls():
    links = []
    for i in urls:
        if i[-1:] == '/':
                i = i[:-1]
        if i not in links and '.onion' in i:
            links.append(i)

    f = open("crawled.txt", "w")
    for url in links:
        f.write(url + '\n')
    f.close()


def setCrawledUrl():
    if os.path.isfile('finished.txt'):
        file = open('finished.txt', "r")
        lines = file.readlines()
        for x in lines:
            crawledUrls.append(x.strip('\n'))


def startSelenium():
    DRIVER_PATH = '/home/linuxlite/tor-browser/Browser/firefox'

    fireFoxOptions = Options()
    fireFoxOptions.binary_location = DRIVER_PATH
    fireFoxOptions.set_preference("browser.download.dir",
                                  "/home/linuxlite/Torloads/")

    fireFoxOptions.set_preference("browser.download.manager.showWhenStarting",
                                  False)

    fireFoxOptions.set_preference("browser.download.folderList", 2)

    fireFoxOptions.set_preference("browser.helperApps.neverAsk.saveToDisk",
                                  "application/octet-stream")

    fireFoxOptions.add_argument('--no-proxy-server')

    driver = webdriver.Firefox(service=Service(GeckoDriverManager().install()),
                               options=fireFoxOptions)

    driver.set_page_load_timeout(60)
    number = randint(33, 43)
    driver.find_element(By.ID, 'connectButton').click()
    print(f'Sleeping␣{number}␣-␣starting␣Tor')
    time.sleep(number)
    print("returning␣driver")
```

```
    return driver


def moveFiles(dest):
    src = "/home/linuxlite/Torloads/"
    for file in os.listdir(src):
        source = src + file
        destination = dest + file
        if os.path.isfile(source):
            shutil.move(source, destination)
            print("Moved␣-␣" + file)
        else:
            print('Not␣moved␣anything')


def loginSelenium(driver):
    driver.find_element(By.NAME, "username").send_keys("StanleyTheCatMan")
    driver.find_element(By.NAME, "password").send_keys("EnToTre123" +
                                                       Keys.ENTER)


if __name__ == "__main__":
    driver = startSelenium()
    setCrawledUrl()
    main(sys.argv[1:], driver)
    print("Programmet␣er␣ferdig!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!")
```

Vedlegg G

Intervju med NC3

Møtereferat fra samtale med NC3.

Møte med Kripos 16.03.23

Deltakere

- Raymond Hagen
- Mathias Viken Borgersen (Kripos)
- Mads Helland Astrup (Kripos)
- Alle gruppemedlemmer

Agenda

- Tips til hvordan man kan gå frem for å finne hackergrupper, type forum, marked, auksjoner?
- Om dere har merket noe til det pågående DDOS angrepet mot Tor nettverket?
- Erfaring med å snakke med Hacker-for-Hire?
- Anbefalte søkemotorer eller søkemotorer som etterforskerne har erfaring med?
- Tanker om bruk av proxy?

Notater

Raymond går gjennom seg selv, og litt bakgrunn for møtet, og kjenner en som heter Simon, som pekte mot Mads Helland Astrup (etterforsker) og Mathias Viken Borgersen (kryptovaluta).

Kripos har i det siste året hatt fokus på misbruk mot barn, og jobbet med å ta ned slik aktivitet.

Ser på det mørke nettet som en del av internettetterforskning, så det går under internett.

Chatrooms med dialoger mellom hackergrupper, auksjoner:

Benytter åpne kilder, Google, for å finne pastebins, og forumer, og auksjoner.
Russian Market, 50 dollar i form av bitcoin.

Altså veldig mye manuelt arbeid.

Er det mye tid som går til etterretningsarbeid:

Mange ulike avdelinger, som samarbeider, mye metodikk for informasjonsinnhenting.

Strengt regulert på hva de kan lagre, og hvor lenge det kan lagres. Riksadvokaten regulerer Kripos.

Utfordring å finne «riktig» informasjon, som brukernavn, markedsplasser. Scraping er en aktuell måte, ressurskrevende måte å scrape selv, mulig outsourcing av programvare av scrapere. Kripos benytter seg av verktøy som de kjøper tilgang til, noe

som er svært kostbart. Bruker disse verktøyene til etterretning, og samle inn data til bruk i operasjoner. Legitime selskaper som selger verktøy.

Holde seg til nordmenn eller utenlandske? Utenlandske

Kripos kan kanskje bidra med linker og lignende, hvis det er av interesse.

Informasjon om hvordan hackergrupper kommuniserer om:

Egen seksjoner som jobber med løspengevirus og skadevare, Konti(?) er en gruppe som er blitt leaket. Grave på twitter og google, mye informasjon og tilgjengelige analyser på Konti(?).

Raymonds bekjente mente at det var 200 stk i Norge som benyttet seg av det mørke nettet.

IP adressen som kobler seg opp mot entry-noden er synlig. Gareth Owenson, har skrevet om Tor og det mørke nettet.

Om Tor blir brukt legitimt senere i tid:

Behovet for anonymitet øker, flere kriminelle ønsker en anonym platform for kommunikasjon. Fler og fler krypterte meldingstjenester og egnede enheter.

Større behov og etterspørsel etter en sikker måte å lagre informasjon.

NyTimes og Facebook var de første legitime sidene på Tor.

Omgåelse av personvern, og myndighetene i seg selv.

Tilgjengelighet som det viktigste i CIA:

Sender angrep gjennom Tor. Folk som driver med Datakrim, arbeider som organisasjoner. Ulike måter å stole på ting, proxy, VPN, Tor, noder. Hvem skal man stole på?

Nysgjerrige folk på Tor – forsvinner fort.

Ofte som politi snakker med annet politi i håp om det er en kriminell.

Mye fokus på overgrep, blir styrt av Riksadvokaten, prioriteres høyere enn f.eks narkotika.

Fremtiden skal datakrim, og cyber prioriteres.

Avsløringsfaktorer: Informasjon, gjenbruk av brukernavn og passord, si for mye av seg selv. Samme IP på nettbank og et annet sted, nok til å avsløre. Bevis forsvinner ved utskifte av enheter. Gjentagende trend ved utskifte av enheter.

Profitt eller en tilfredstillelse. Mange måter å skjule seg selv, men mange måter å gjøre feil på, menneskelige feil. Kryptosporing, sporing av personinformasjon, ved video, bilder eller tekst.

Teknisk dyktige folk blir late, glemmer å gjøre ting. Mange vurderinger, vil føre til en feilvurdering.

Veldig mange som følger med, gjennom både private og offentlige sektorer, og internasjonale. Så mange som følger med på aktivitet.

Har dere noen tips til hvordan man kan gå frem for å finne hackergrupper, type forum, marked, auksjoner?

Kripos har lik problemstilling, og avhenger av manuell etterforskning. Telegram-grupper og ulike forum, benytter egen programvare for indeksering, slik som de kan søke gjennom. Følger trender som er aktuelle for dagens problemer.

Om dere har merket noe til det pågående DDOS angrepet mot Tor nettverket?

Registrert, men ikke noe særlig relevant for Kripos da det må være en sak i grunn.

Erfaring med å snakke med Hacker-for-Hire?

Driver med digital etterforskning, for innhenting av informasjon.

Søkemotorer dere anbefaler / har erfaring med?

Mye scrapeaktivitet. Kan ikke kommentere hvilke
Ahmia, torch

Sliter med captcha, så de har mennesker som løser disse manuelt.

Proxy

Hvis man skal være en viss person, så funker det bra. Gjelder i bunn og grunn hvem man stoler på. Vurdering på pris og tid. Objektivt sett er det tryggere. Kanskje man skal fremstå som en person som er kapabel.
VPN over proxy.

Vedlegg H

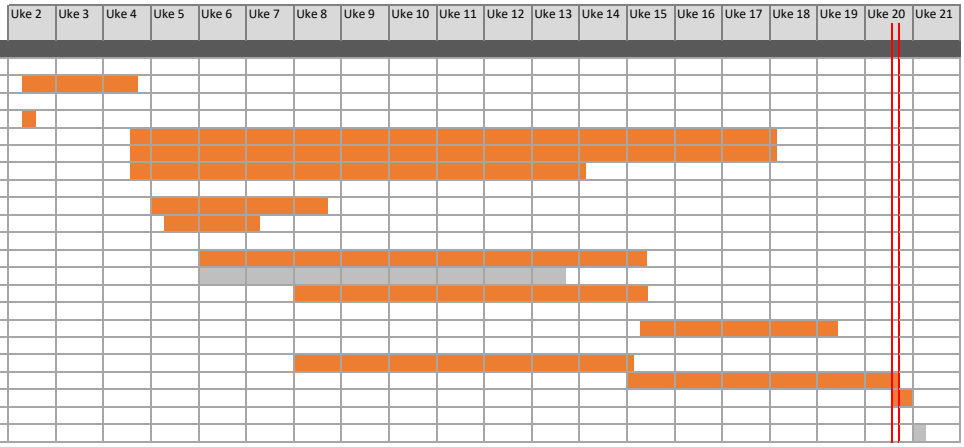
Gantt-skjema

Dette er det oppdaterte Gantt-skjemaet.

Høsting fra det mørke nettet

Prosjekt start: ons, 11.jan.2023
Se uke: 1

OPPGAVE	FREMGANG	START	SLUTT
Fase 1: Prosjektplan			
1.1 Prosjektplan	100 %	11.jan	27.jan
Fase 2: Informasjonshenting			
2.1 Oppsett av PC'er	100 %	11.jan	12.jan
2.2 Bakgrunn og historie av Dark Web	100 %	27.jan	01.mai
2.3 Teknisk oppbygging av Dark Web	100 %	27.jan	01.mai
2.4 Navigasjon inne i Dark Web	100 %	27.jan	03.apr
Fase 3: Design av crawler			
3.1 Informasjonshenting (eksisterende crawler-koder)	100 %	30.jan	24.feb
3.2 Oppsett og programmeringsspråk	100 %	01.feb	14.feb
Fase 4: Crawler			
4.1 Utvikling av crawler	100 %	06.feb	12.apr
4.2 Utvikling av database med API	0 %	06.feb	31.mar
4.3 Testing av crawler	100 %	20.feb	12.apr
Fase 5: Kjøring og høsting			
5.1 Kjøre crawleren	100 %	12.apr	10.mai
Fase 6: Rapportskriving			
6.1 Lage utkast av rapport til veileder	100 %	20.feb	10.apr
6.2 Kun skrive rapport	100 %	10.apr	19.mai
6.3 Finpuss og ferdigstilling av rapport	100 %	19.mai	21.mai
Fase 7: Overlevering av samlet data			
7.1 Overlevering av database med samlet data	0 %	22.mai	23.mai



Vedlegg I

Timelister

Marte

Uke	Dag	Antall timer	Beskrivelse	Total
2	Mandag			11
	Tirsdag	1	Statusmøte	
	Onsdag	7	Projektplan og seminar	
	Torsdag			
	Fredag	3	Prosjektplanarbeid	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
3	Mandag	7	Arbeid med projektplan	23
	Tirsdag	3	Arbeid med prosjektplan	
	Onsdag	6	Arbeid med projektplan	
	Torsdag			
	Fredag	7	Arbeid med projektplan + møte med veileder og oppdragsgiver	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
4	Mandag	6	Ferdigstille prosjektplan	18
	Tirsdag			
	Onsdag	2	Ferdigstille prosjektplan	
	Torsdag	5	Møte veileder og ferdigstille prosjektplan	
	Fredag	5	Møte oppdragsgiver og ferdigstille prosjektplan	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
5	Mandag	3	Research Tor	16
	Tirsdag	2	Research generelt om Tor	
	Onsdag			
	Torsdag	6	Research bakgrunn/historie Tor	
	Fredag			
	Lørdag	5	Leite etter litteratur	
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
6	Mandag	6	Bakgrunn og historie Tor	26
	Tirsdag			
	Onsdag	7	Forskning/litteratur Tor	
	Torsdag	8	Forskning/litteratur Tor	
	Fredag	5	Forskning/litteratur Tor	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
7	Mandag	8	Bakgrunn og historie	22
	Tirsdag	6	Research	
	Onsdag			
	Torsdag			
	Fredag	8	Research + møte med Raymond	

	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
8	Mandag	8	Bakgrunn research	29
	Tirsdag			
	Onsdag	4	Møte oppdragsgiver + research + Tor	
	Torsdag			
	Fredag	8	Navigere og lete lenker på Dark Web	
	Lørdag	4	Navigere og lete lenker på Dark Web	
	Søndag	5	Navigere og lete lenker på Dark Web	
Uke	Dag	Antall timer	Beskrivelse	Total
9	Mandag			28
	Tirsdag	6	Lynkurs 2 + rapportskriving	
	Onsdag			
	Torsdag	8	Skrive rapport	
	Fredag	9	Skrive rapport og møte med oppdragsgiver	
	Lørdag	5	Rapportskriving	
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
10	Mandag	7	Oppgavestruktur, overskrifter, spørreus	29
	Tirsdag	4	Rapportskriving	
	Onsdag	5	Rapportskriving	
	Torsdag			
	Fredag	7	Gå gjennom tilbakemeldinger oppdragsgiver + mer skriving	
	Lørdag	6	Rapportskriving	
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
11	Mandag	8	Rapportskriving	32
	Tirsdag	8	Rapportskriving	
	Onsdag			
	Torsdag	9	Rapportskriving + møte med Kripos	
	Fredag			
	Lørdag	7	Rapportskriving	
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
12	Mandag	7	Rapportskriving	27
	Tirsdag	8	Rapportskriving	
	Onsdag			
	Torsdag	5	Møte med veileder + rapportskriving	
	Fredag	7	Rapportskriving	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
	Mandag	4	Navigere og lete lenker på forum	
	Tirsdag	5	Navigere på Tor	
	Onsdag	3	Navigere Tor + åpen nett	

13	Torsdag			21
	Fredag	9	Rapportskriving + møte med oppdragsgiver	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
14	Mandag	5	Rapportskriving	24
	Tirsdag	4	Rapportskriving	
	Onsdag	7	Rapportskriving	
	Torsdag	3	Rapportskriving	
	Fredag			
	Lørdag	5		
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
15	Mandag	6	Rapportskriving	33
	Tirsdag	9	Rapportskriving	
	Onsdag	8	Rapportskriving	
	Torsdag	4	Rapportskriving	
	Fredag			
	Lørdag	6	Rapportskriving	
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
16	Mandag	9	Rapportskriving	48
	Tirsdag	6	Rapportskriving	
	Onsdag	6	Rapportskriving	
	Torsdag	7	Rapportskriving	
	Fredag	8	Rapportskriving + møte med veileder og oppdragsgiver	
	Lørdag	7	Rapportskriving + jobbe med tilbakemeldinger	
	Søndag	5	Rapportskriving - utkast	
Uke	Dag	Antall timer	Beskrivelse	Total
17	Mandag	7	Rapportskriving - utkast	42
	Tirsdag	5	Rapportskriving - utkast	
	Onsdag	10	Rapportskriving - utkast	
	Torsdag	7	Rapportskriving	
	Fredag	8	Rapportskriving	
	Lørdag			
	Søndag	5	Rapportskriving	
Uke	Dag	Antall timer	Beskrivelse	Total
18	Mandag	7	Rapportskriving	49
	Tirsdag	5	Rapportskriving	
	Onsdag	11	Rapportskriving	
	Torsdag	6	Rapportskriving	
	Fredag	8	Rapportskriving, møte med oppdragsgiver	
	Lørdag	7	Rapportskriving	
	Søndag	5	Rapportskriving	
Uke	Dag	Antall timer	Beskrivelse	Total

19	Mandag	9	Rapportskriving	57
	Tirsdag	6	Rapportskriving	
	Onsdag	7	Rapportskriving	
	Torsdag	9	Rapportskriving	
	Fredag	7	Rapportskriving	
	Lørdag	9	Rapportskriving	
	Søndag	10	Rapportskriving	
Uke	Dag	Antall timer	Beskrivelse	Total
20	Mandag	8	Rapportskriving	54
	Tirsdag	10	Rapportskriving	
	Onsdag			
	Torsdag	10	Rapportskriving	
	Fredag	10	Rapportskriving	
	Lørdag	8	Finnpuss	
	Søndag	8	Finpuss	

Martin

Uke	Dag	Antall timer	Beskrivelse	Total
2	Mandag			13
	Tirsdag	1	Statusmøte	
	Onsdag	7	Prosjektplan og lynkurs	
	Torsdag	2	Studie av tidligere oppgaver	
	Fredag	3	Prosjektplanarbeid	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
3	Mandag	7	Arbeid med prosjektplan	25
	Tirsdag	3	Gantt-skjema draft	
	Onsdag	5	Arbeid med prosjektplan	
	Torsdag	3	Gantt-skjema ferdigstilling	
	Fredag	5	Møter med veileder og oppdragsgiver, arbeid med prosjektplan	
	Lørdag	2	Fikla med Gantt-skjema inn i prosjektplanen	
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
4	Mandag	7	Prosjektplanarbeid	24
	Tirsdag	5	Prosjektplanarbeid	
	Onsdag	4	Gjorde research på det mørke nettet	
	Torsdag	4	Prosjektplanarbeid	
	Fredag	4	Møte oppdragsgiver og ferdigstille prosjektplan	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
5	Mandag	6	Ferdigstilling og levering av prosjektplanen	28
	Tirsdag	5	Undersøke crawlere og kodespråk	
	Onsdag	6	Undersøke crawlere og kodespråk	
	Torsdag	5	Statusmøte	
	Fredag	6	Statusmøte med veileder, undersøke kodebibliotek	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
6	Mandag	7	Startet med crawleren	30
	Tirsdag	6	Utvikling av crawler	
	Onsdag	6	Utvikling av crawler	
	Torsdag	6	Møte med veileder og interntmøte	
	Fredag	5	Møte med oppdragsgiver og arbeid med crawleren	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total

7	Mandag	6	Utvikling av crawler	31
	Tirsdag	7	Utvikling av crawler	
	Onsdag	5	Utvikling av crawler	
	Torsdag	6,5	Utvikling av crawler	
	Fredag	6,5	Utvikling av crawler og møte med oppdragsgiver	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
8	Mandag	6	Startet med skriving av metode av crawleren	30
	Tirsdag	6	Startet med skriving av metode av crawleren	
	Onsdag	7	Møte med oppdragsgiver, utvikling av crawler	
	Torsdag	6	Møte med veileder og skriving av rapport	
	Fredag	5	Utvikling av crawler	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
9	Mandag	7	Utvikling av crawler	32
	Tirsdag	5	Lynkurs, rapportskriving	
	Onsdag	7	Skriving på rapport	
	Torsdag	6	Planlegging og laging av flytdiagram	
	Fredag	7	Skriving på rapport og møte med oppdragsgiver	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
10	Mandag	8	Lage skjelett til rapporten, og spørreundersøkelse	34
	Tirsdag	6	Arbeid rundt rapport	
	Onsdag	7	Laget nytt flytdiagram og link-research	
	Torsdag	7	Skriving på rapport, og klassemøte	
	Fredag	6	Rapportskriving og møte med oppdragsgiver	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
11	Mandag	6,5	Rapportskriving og utvikling av crawler	31
	Tirsdag	6	Utvikling av crawler	
	Onsdag	7	Rapportskriving og utvikling av crawler	
	Torsdag	6,5	Utvikling av crawler og møte med Kripos	
	Fredag	5	Utvikling av crawler og møte med oppdragsgiver	
	Lørdag			
	Søndag			

Uke	Dag	Antall timer	Beskrivelse	Total
12	Mandag	5	Rapportskriving	29,5
	Tirsdag	6	Utvikling crawler og rapportskriving	
	Onsdag	7	Utvikling crawler og rapportskriving	
	Torsdag	5,5	Møte med veileder, og utvikling av crawler	
	Fredag	6	Utvikling crawler og leting etter urler	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
13	Mandag	6	Utvikling av crawler og rapportskriving	24
	Tirsdag	5	Utvikling av crawler og rapportskriving	
	Onsdag	4	Rapportskriving	
	Torsdag	4	Rapportskriving	
	Fredag	5	Arbeid med 1.utkast til veileder	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
14	Mandag	6	Arbeid med 1.utkast til veileder	22
	Tirsdag	5	Arbeid med 1.utkast til veileder	
	Onsdag	5	Arbeid med 1.utkast til veileder	
	Torsdag			
	Fredag			
	Lørdag			
	Søndag	6	Utvikling av crawler	
Uke	Dag	Antall timer	Beskrivelse	Total
15	Mandag			30
	Tirsdag	7	Rapportskriving	
	Onsdag	6	Rapportskriving	
	Torsdag	6	Rapportskriving	
	Fredag	7	Rapportskriving og møte med oppdragsgiver + funn av noen vulkan-filer	
	Lørdag	4	Litt rapportskriving og kildeleting	
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
16	Mandag	5	Rapportskriving	29,5
	Tirsdag	4,5	Rapportskriving og rekognisering av dark web domener	
	Onsdag	6	Rapportskriving	
	Torsdag	7	Rapportskriving	
	Fredag	7	Rapportskriving	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
	Mandag	7	Rapportskriving	
	Tirsdag	5	Rapportskriving	
	Onsdag	6	Rapportskriving	

17	Torsdag	6	Rapportskriving	31
	Fredag	7	Rapportskriving og møte med oppdragsgiver	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
18	Mandag	7	Rapportskriving	42
	Tirsdag	8	Rapportskriving	
	Onsdag	8	Rapportskriving	
	Torsdag	6	Rapportskriving	
	Fredag	7	Rapportskriving og møte med oppdragsgiver	
	Lørdag	6	Rapportskriving	
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
19	Mandag	6	Rapportskriving	48
	Tirsdag	8	Rapportskriving	
	Onsdag	8	Rapportskriving	
	Torsdag	8	Rapportskriving	
	Fredag	7	Rapportskriving	
	Lørdag	6	Rapportskriving	
	Søndag	5	Rapportskriving	
Uke	Dag	Antall timer	Beskrivelse	Total
20	Mandag	9	Rapportskriving	57
	Tirsdag	10	Rapportskriving	
	Onsdag	4	Rapportskriving	
	Torsdag	10	Rapportskriving	
	Fredag	8	Rapportskriving	
	Lørdag	8	Finpuss ++	
	Søndag	8	Finpuss	

Mats

Uke	Dag	Antall timer	Beskrivelse	Total
2	Mandag			13
	Tirsdag	1	Statusmøte	
	Onsdag	7	Projektplan og seminar	
	Torsdag	2	Lese gjennom tidligere oppgaver	
	Fredag	3	Prosjektplanarbeid	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
3	Mandag	7	Arbeid med projektplan	25
	Tirsdag	3	Gantt-skjema draft	
	Onsdag	6	Arbeid med projektplan	
	Torsdag	3	Arbeid med projektplan	
	Fredag	6	Arbeid med projektplan + møte med veileder og oppdragsgiver	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
4	Mandag	6	Ferdigstille prosjektplanen	26
	Tirsdag	3	Satte opp EkspressVPN	
	Onsdag	5	Research på det mørke nettet	
	Torsdag	5	Møte veileder og jobbe med prosjektplan	
	Fredag	7	Møte oppdragsgiver og jobbe med prosjektplan	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
5	Mandag	6	Ferdigstille og levere prosjektplan	29
	Tirsdag	5	Undersøke crawlere og kodespråk	
	Onsdag	7	Undersøke crawlere og kodespråk	
	Torsdag	6	Statusmøte	
	Fredag	5	statusmøte med oppgavegiver og undersøke bibliotek til crawlere	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
6	Mandag	7	Startet med crawler	32
	Tirsdag	5	Utvikling	
	Onsdag	6	Undersøke bibliotek og lignende kode	
	Torsdag	6	Møte med veileder og internt	
	Fredag	8	møte med oppgavegiver, og jobbing md crawler	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
	Mandag	7	Jobbing med crawler	
	Tirsdag	7	Fikset RaspAP - VPN	

7	Onsdag	6	Utvikling crawler	28
	Torsdag	3,5	Utvikling crawler	
	Fredag	4,5	Testkjørt crawler på dark web	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
8	Mandag	6	Utvikling crawler og rapportskriving	34
	Tirsdag	4	Jobbing med rapport	
	Onsdag	7	Utvikling crawler	
	Torsdag	6	Testing og utvikling	
	Fredag	8	Utvikling crawler	
	Lørdag	3	Utvikling crawler	
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
9	Mandag	6	Utvikling crawler	27
	Tirsdag	5	Utvikling crawler	
	Onsdag	7	Utvikling crawler og rapportskriving	
	Torsdag	4	Navigering og utvikling	
	Fredag	5	Utvikling crawler	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
10	Mandag	6	Lage skjelett til rapport, og spørreundersøkelse	25
	Tirsdag	5	Jobbe med rapport	
	Onsdag	5	Lage flytdiagram og leting etter urler	
	Torsdag	5	Navigering og rapport	
	Fredag	4	Statusmøte med veileder	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
11	Mandag	6	Rapportskriving	37
	Tirsdag	6	Utvikling crawler	
	Onsdag	7	Utvikling crawler og rapportskriving	
	Torsdag	6	Møte med kripos og utvikling av crawler	
	Fredag	5	Flytdiagram og rapportskriving	
	Lørdag	4	Rapportskriving	
	Søndag	3	Rapportskriving	
Uke	Dag	Antall timer	Beskrivelse	Total
12	Mandag	5	Rapportskriving	30
	Tirsdag	7	Utvikling crawler og rapportskriving	
	Onsdag	5	Utvikling crawler og rapportskriving	
	Torsdag	4	Navigering og testing	
	Fredag	6	Utvikling crawler og leting etter urler	
	Lørdag			
	Søndag	3	Forbedringer på VPN	
Uke	Dag	Antall timer	Beskrivelse	Total
	Mandag	6	Utvikling crawler og rapportskriving	

13	Tirsdag	3	Utvikling crawler	30
	Onsdag	5	Navigering og rapport	
	Torsdag	4	Rapportskriving	
	Fredag	5	Rapportskriving	
	Lørdag	4	Arbeid med 1.utkast til veileder	
	Søndag	3	Arbeid med 1.utkast til veileder	
Uke	Dag	Antall timer	Beskrivelse	Total
14	Mandag	6	Arbeid med 1.utkast til veileder	21
	Tirsdag	4	Arbeid med 1.utkast til veileder	
	Onsdag	7	Arbeid med 1.utkast til veileder	
	Torsdag			
	Fredag			
	Lørdag			
	Søndag	4	Utvikling crawler	
Uke	Dag	Antall timer	Beskrivelse	Total
15	Mandag	8	Utvikling crawler og rapportskriving	39
	Tirsdag	7	Utvikling crawler og rapportskriving	
	Onsdag	5	Oppstart av crawler på 3x maskiner + manuelt arbeid	
	Torsdag	7	Rapportskriving og skraping	
	Fredag	7	Rapportskriving og skraping	
	Lørdag	3	Skraping	
	Søndag	2	Skraping	
Uke	Dag	Antall timer	Beskrivelse	Total
16	Mandag	7	Rapportskriving og skraping	43
	Tirsdag	6	Rapportskriving og skraping	
	Onsdag	8	Rapportskriving og skraping	
	Torsdag	8	Rapportskriving og skraping	
	Fredag	8	Rapportskriving og skraping	
	Lørdag	3	Skraping	
	Søndag	3	Skraping	
Uke	Dag	Antall timer	Beskrivelse	Total
17	Mandag	7	Rapportskriving og skraping	39
	Tirsdag	8	Rapportskriving og skraping	
	Onsdag	8	Rapportskriving og skraping	
	Torsdag	8	Rapportskriving og skraping	
	Fredag	6	Rapportskriving og skraping	
	Lørdag	1	Skraping	
	Søndag	1	Skraping	
Uke	Dag	Antall timer	Beskrivelse	Total
18	Mandag	7	Rapportskriving	43
	Tirsdag	5	Rapportskriving	
	Onsdag	9	Rapportskriving	
	Torsdag	6	Rapportskriving	
	Fredag	8	Rapportskriving og møte med oppdragsgiver	
	Lørdag	3	Rapportskriving og skraping	
	Søndag	5	Rapportskriving og skraping	

Uke	Dag	Antall timer	Beskrivelse	Total
19	Mandag	6	Rapportskriving og skraping	50
	Tirsdag	9	Rapportskriving og skraping	
	Onsdag	8	Rapportskriving og aller siste skraping	
	Torsdag	8	Rapportskriving og fikse data fra skraping	
	Fredag	8	Rapportskriving	
	Lørdag	6	Rapportskriving	
	Søndag	5	Rapportskriving	
Uke	Dag	Antall timer	Beskrivelse	Total
20	Mandag	8	Rapportskriving	51
	Tirsdag	7	Rapportskriving	
	Onsdag			
	Torsdag	10	Rapportskriving	
	Fredag	10	Rapportskriving	
	Lørdag	8	Finpuss rapport	
	Søndag	8	Finpuss rapport	

Uke	Dag	Antall timer	Beskrivelse	Total
2	Mandag			13
	Tirsdag	1	Statusmøte	
	Onsdag	7	Projektplan og seminar	
	Torsdag	2	Lese eksempeloppgave	
	Fredag	3	Prosjektplanarbeid	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
3	Mandag	7	Arbeid med projektplan	25
	Tirsdag	3	Timelister og litt prosjektplanarbeid	
	Onsdag	6	Arbeid med projektplan	
	Torsdag	3	Arbeid med projektplan	
	Fredag	6	Arbeid med projektplan	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
4	Mandag	7	Prosjektplanarbeid	25
	Tirsdag	1	Prosjektplanarbeid	
	Onsdag	4	Research	
	Torsdag	6	Møte med veileder + prosjektplanarbeid	
	Fredag	7	Møte med oppdragsgiver + prosjektplan	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
5	Mandag	6	Ferdigstilling av prosjektplan og research	28
	Tirsdag	1	Research teknisk oppbygging	
	Onsdag	6	Navigasjon på Tor ++	
	Torsdag	7	Statusmøte med veileder og research	
	Fredag	8	Statusmøte med oppdragsgiver + research	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
6	Mandag	7	Research teknisk oppbygging	22
	Tirsdag	2	Navigasjon + research teknisk oppbygging	
	Onsdag	2	Navigasjon/ research	
	Torsdag	6	Møte med Erjon(veileder) og research	
	Fredag	5	Litt research og møte med oppdragsgiver	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
7	Mandag	6	Litteraturstudie	25
	Tirsdag	2	Rapportskriving	
	Onsdag			
	Torsdag	7	Litteraturstudie	
	Fredag	6	Litteraturstudie	
	Lørdag			
	Søndag	4	Navigasjon på Tor ++	

Thugitha

Uke	Dag	Antall timer	Beskrivelse	Total
8	Mandag	6	Rapportskriving	25
	Tirsdag	2	Litteraturstudie	
	Onsdag	7	Rapportskriving + møte med oppdragsgiver	
	Torsdag	4	Litteraturstudie + møte med veileder	
	Fredag	6	Rapportskriving	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
9	Mandag	4	Litt rapportskriving + research	25
	Tirsdag	4	Seminar med Frode + rapportskriving	
	Onsdag			
	Torsdag	7	Litteraturstudie	
	Fredag	6	Litteraturstudie	
	Lørdag	4	Utforsking av Tor og div onion-sites	
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
10	Mandag	5	Rapportskriving	19
	Tirsdag			
	Onsdag	5	Litteraturstudie	
	Torsdag	4	Rapportskriving	
	Fredag	5	Rapportskriving + møte med oppdragsgiver	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
11	Mandag	6	Rapportskriving	29
	Tirsdag	3	Rapportskriving	
	Onsdag	6	Litteraturstudie	
	Torsdag	7	Litteraturstudie + møte med kripos	
	Fredag	7	Rapportskriving + møte med oppdragsgiver	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
12	Mandag	5	Navigasjon på Tor for å finne onion-sites	24
	Tirsdag	3	Rapportskriving	
	Onsdag	5	Rapportskriving	
	Torsdag	6	Rapportskriving	
	Fredag	5	Retting av småfeil + struktur	
	Lørdag			
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
13	Mandag	6	Litteraturstudie	29
	Tirsdag	4	Litteraturstudie	
	Onsdag	5	Rapportskriving	
	Torsdag	6	Rapportskriving	
	Fredag	8	Rapport+statusmøte med oppdragsgiver	
	Lørdag			
	Søndag			

Uke	Dag	Antall timer	Beskrivelse	Total
14	Mandag	6	Arbeid med 1.utkast til veileder	34
	Tirsdag	4	Arbeid med 1.utkast til veileder	
	Onsdag	7	Arbeid med 1.utkast til veileder	
	Torsdag	6	Arbeid med 1.utkast til veileder	
	Fredag	7	Arbeid med 1.utkast til veileder	
	Lørdag			
	Søndag	4	Rapporskriving	
Uke	Dag	Antall timer	Beskrivelse	Total
15	Mandag	7	Rapportskriving	39
	Tirsdag	5	Rapportskriving	
	Onsdag	8	Rapportskriving	
	Torsdag	7	Rapportskriving	
	Fredag	8	rapport + møte med oppdragsgiver	
	Lørdag			
	Søndag	4	Rapporskriving	
Uke	Dag	Antall timer	Beskrivelse	Total
16	Mandag	7	Rapportskriving	43
	Tirsdag	9	Rapportskriving	
	Onsdag	9	Rapportskriving	
	Torsdag	8	Rapportskriving	
	Fredag	7	rapport + møte med oppdragsgiver + veiled	
	Lørdag			
	Søndag	3	Rapportskriving	
Uke	Dag	Antall timer	Beskrivelse	Total
17	Mandag	8	Arbeid med 2.utkast til veileder	41
	Tirsdag	5	Arbeid med 2.utkast til veileder	
	Onsdag	8	Arbeid med 2.utkast til veileder	
	Torsdag	9	Jobbet videre med rapporten	
	Fredag	9	Møte med oppdragsgiver + egenarbeid	
	Lørdag	2	småpirk på rapport	
	Søndag			
Uke	Dag	Antall timer	Beskrivelse	Total
18	Mandag	8	Rapportskriving	30
	Tirsdag	9	Møte med veileder + forbedring av rapport	
	Onsdag	9	Rapportskriving	
	Torsdag	2	Rapportskriving	
	Fredag		Helgetur til Amsterdam	
	Lørdag		Helgetur til Amsterdam	
	Søndag	2	småpirk på rapport	
Uke	Dag	Antall timer	Beskrivelse	Total
19	Mandag	9	Rapportskriving	53
	Tirsdag	8	Rapportskriving	
	Onsdag	9	Rapportskriving	
	Torsdag	9	Rapportskriving	
	Fredag	8	Rapportskriving	
	Lørdag			
	Søndag	10	Rapportskriving	

Uke	Dag	Antall timer	Beskrivelse	Total
20	Mandag	9	Rapportskriving	63
	Tirsdag	10	Rapportskriving	
	Onsdag	7	Rapportskriving	
	Torsdag	10	Rapporskriving	
	Fredag	11	Rapportskriving	
	Lørdag	8	Finpuss + omskriving	
	Søndag	8	Finpuss	

Vedlegg J

Møtereferater fra møter med oppdragsgiver

Alle møtereferatene fra møter med oppdragsgiver.

09.01.23 – Oppstartsmøte med oppdragsgiver

Deltakere

- Raymond Hagen (oppdragsgiver)
- Alle gruppemedlemmer

Agenda

- Diskutere fremdriftsplan
- Avtale faste møtetidspunkter
- Avklare mål med oppgaven

Fremdriftsplan

- Få en oversikt over hva som må gjøres og sette oss delmål, blant annet i form av Gantt-skjema og prosjektplan

Faste møter

- Det vil bli holdt faste møter med oppdragsgiver fredag kl.13 for å oppdatere og sørge for god progresjon

Mål

- Læringsutbytte; opparbeide oss kunnskap om det mørke nettet
- Høsting av data fra det mørke nettet over i en database

20.01.23 – Statusmøte med oppdragsgiver

Deltakere

- Raymond Hagen (oppdragsgiver)
- Marte Jørgensen
- Martin Hyldmo
- Mats Dimitri Jensen

Agenda/spørsmål

- Mål og forventninger
- Politiet
- Vi fikk ikke tak i VPN, er dette noe du kan fikse?
- Skal vi se på dataen før vi overleverer den?
- Vil du ha alt av innhold fra hver side f.eks. bilder, videoer osv.? Eller kun tekst/html
- Har du minnepenner/harddisker vi kan bruke?
- Har du noen spesifikke ord vi kan gå etter for å finne riktig type data?
- Har du noen søkemotorer du anbefaler for dark web?
- Hvordan ser du på risikoen ved å kjøre crawleren på hjemmenettet? (med VPN)
- Hvordan er du når det kommer til koding? Kan vi bruke deg dersom vi står fast?
- Informasjon om forskningen til Raymond.

Mål og forventninger

- Først snakker vi litt om prosjektplanen, hva vi har tenkt og får input i hva Raymond tenker er viktig.
- Det viktige er at vi klarer å legge på våre egne vurderinger, ikke bare gjengi det som står allerede.
- Omfang: Sette problemstilling: Noe om anonymisering, protokoll -> hensikt -> gir visse muligheter, både bra og dårlig.
- Det er godt dokumentert at det som foregår i hackermiljø, eksponering, deling av tips, koordinering av operasjoner osv. blir det mørke nettet brukt til. Målet er da å prøve å hente mest mulig data ifht den problemstillingen(!) ikke alt anna i tillegg.
    - Innblikk, forståelse til hvordan de opererer
- Fortelle mer om bakgrunnen til det mørke nettet og TOR i seg selv i prosjektplan? Kan være lurt.
- Viktig å få med historien!! Begynte som stuerent, men ble misbrukt og der er vi i dag.
- Noen åpenbare punkt: Må ha en start, hvor er de rette nettsidene, hvordan finner du ut hvor du skal lete?
    - Diskutere ulemper med det mørke nettet: Det er tregt, høy brukerterskel, skalerer ikke, men anonymiteten f.eks. veier opp?

Politiet

- Politiet: Sannsynligheten for at vi finner noe som er ulovlig er relativt høy, men vi har fått en godkjenning fra NTNU og DigDir - forskning. Det er klart at hvis vi kommer over noe ifht barn - ring kripos!
- Raymond er ikke så bekymret for dette, da også politiet vet at det finnes forskere i tillegg til kriminelle der. Alt blir overvåket.
- Scraper tilfeldigvis laster ned noe ulovlig, prøve å stoppe det, ikke laste ned
- Hvis vi er redd for å finne barnepornografi, kan prøve å finne TOR-noder i Norge er det mest sannsynlig filtrert vekk.
- Finner dokumenter, credentials, ransomware-data ok å laste ned.
- Etisk dilemma?

VPN

- Dark web på skolenett: Usikkert, spørre noen andre. Vi tenker at vi spør IT avdelingen på NTNU.
- VPN: Får ikke fra NTNU. Får ikke fra han, så da forslår han at vi bruker VPN til skolen. Gjestenett?

Data overlevering

- Se gjennom data: Tja, nei, det kan være ekstremt mye data så egentlig ikke. Vi kan bestemme selv.
- Data: Alt av innhold, bilder og video – Ja takk alt

Harddisk

- Raymond fikser - kommer i midten i februar

Nøkkelord for søk

- Kommer tilbake til det, skal tenke på det. "whale ops" er ett eksempel

Søkemotor

- DuckDuckGo. The hidden wiki er lurest å starte på (adressen endrer seg hele tiden, så må lete litt).

Risiko, crawler på hjemmenett

- Det går fint, så lenge det er med VPN! Nord VPN f.eks.

Veiledning for oppbygning av crawler

- Crawler: Systematisering: Bruker allerede lagde crawlere, ulempen at de kriminelle beskytter seg mot det. Så finne lure løsninger for å omgå dette, leke oss for å finne triks for å finne relevant data.
- Praktisk: Må lete seg igjennom og finner steder som vi kan bruke som input til utviklingen av crawler: utfordring!
    - Dette tar tid, men dokumenter underveis. Vær systematisk, hvor begynner vi jakten? Verktøy, chatroom, snakker vi med folk osv.
- Koding: Raymond koder lite, men kan alltid sende kode og spørre om hjelp om vi står helt fast.

Mer informasjon om forskningen til Raymond:

Mye er strukturert, lett å gå inn i en brannvegg, nettverkskomponent, hva skjedde når vi ble angrepet. Analyseres. Greit nok. Problemet er at en menneskelig intensjon, folk deler meninger på nett, Twitter, blogg, dark web - ustrukturert. Jakter etter strukturerte data, noen miljøer som snakker veldig mye om f.eks. norske gassinstallasjoner/virksomheter. Spørsmålet er kan dette kombinert med krigen Ukraina og andre ting i samfunnet, er det en indikasjon mot f.eks. Equinor? Finne sammenhenger og mønstre.

Finne bin-sites der det ligger masse stjele data fra en bransje eller visse type firmaer som gir indikasjon på at her er det noe som er mer interessant enn andre på visse tidspunkt. Kan brukes til å tenke her kan jeg se på dataene at jeg kan korrelere det til en eller annen hendelse som har skjedd, evt her kan det være indikasjon på at noe kan skje

Matematikk for å gjøre en form for predikasjon om hva som kommer til å skje for å forhindre angrep.

Hackergrupper bruker det mørke nettet for å dele og kommunisere, tror R. Om det stemmer har man en god begynnelse for å et startpunkt i hvor å leite etter infoen man ønsker.

- "Hiding in plain sight

27.01.23 – Statusmøte med oppdragsgiver

Deltakere

- Raymond Hagen (oppdragsgiver)
- Alle gruppemedlemmer

Agenda/spørsmål

- Rammer for oppgaven vår
- Språk vi skal hente data på
- Annet

Rammer

- Få en forståelse for hvordan det mørke nettet er bygd opp, hva protokollen «The onion routing protocol» gjør, og hva den eventuelt ikke gjør. I etterkant av møtet har vi funnet ut at dette går mer under problemområde.
- Lage en crawler som kryper rundt på det mørke nettet og høster inn data. Får man til dette med automatikk, eller må det blandes inn menneskelig faktor? Dette går også mer under problemområdet.

Språk

- Vil ha i alle mulige språk. Har programvarer for å finne ut språk og oversette dette til ønsket språk.
- I nyere tid har forskjellige statlige aktører kommunisert og lagd programmer i et annet språk, for eksempel Kina skriver på russisk og Russland skriver på kinesisk.

Annet

- Interessert i steder hvor hackere typisk poster eller kommuniserer, for eksempel binsites eller forum.
- Hvis man finner relevant data bør dette skrapes ned med en gang, da det ikke er sikkert at denne siden fortsatt er oppe bare timer senere.
- Vil gjerne ha inn gamle virus
- Dark web er veldig overvåket, og selv om data gjerne blir liggende der, så blir det flyttet fra sted til sted.
- Gjerne hold en oversikt om sider som har vært oppe, men som har blitt tatt ned i senere tid.
- Dersom det ikke blir funnet noe relevant data, kan dette også bli sett på som et funn
- Gjerne prøvekjør crawleren på NY times sine sider, eller andre typer sider der du kan sammenligne data, og se hvor effektivt den er.
- Bruk Anonymous sine sider til å finne data om tidligere operasjoner/angrep, dette vil gi en viss indikasjon på hvordan man kan spore andre type grupper.
- Bruk kun egne pc-er til å dokumentere, og kun de utgitte maskinene til leting etter statlige hackergrupper på det mørke nettet.

10.02.23 – Statusmøte med oppdragsgiver

Deltakere

- Raymond Hagen (oppdragsgiver)
- Martin Hyldmo
- Mats Dimitri Jensen
- Thugitha Kanavathi

Agenda

- Statusoppdatering
- Spørsmål om crawler
- Spørsmål

Statusoppdatering

- Begynt å utvikle crawleren og testet den på det åpne nettet
- Fortsetter med research (bakgrunn og teknisk oppbygging)

Spørsmål om crawler

- Crawle først også scrape eller gjøre begge deler samtidig?

Oppdragsgiver ønsker at dette skal foregå samtidig da dette vil resultere i høsting av en større mengde at data.
Han har også tipset oss om å finne binsites da dette er dark webs versjon av pastebin og kan inneholde mye nyttig informasjon.

Annet

- Avtalte fysisk møte uke 8, oppdragsgiver er i Gjøvik

22.02.23 – Statusmøte med oppdragsgiver

Deltakere

- Raymond Hagen (oppdragsgiver)
- Martin Hyldmo
- Mats Dimitri Jensen
- Marte Jørgensen
- Thugitha Kanavathi

Agenda

- Statusoppdatering
- Råd og veiledning

Statusoppdatering

- Prøvd å kjøre crawleren på det mørkenettet
    - Funker som det skal

Råd og veiledning

- Fokuser på at det mørke nettet er et verktøy
    - Laget med en fornuftig baktanke
    - Hva kan det mørke nettet brukes til (som det ikke brukes til allerede)?
        - Legal bruk: verne vanlige brukere mot cookies
- Flowchart for identifisering av område
    - prosedyre
    - Søkemotorer
    - Hvordan går vi frem for å finne nettsider å scrape
- Hva Raymond vil få ut av oppgaven: finne hacker gruppes operasjonelle plan og informasjon om de
- Hva vi ønsker å få ut av oppgaven: Forstå mørke nettet og finne informasjon
- Tittel på hovedrapporten: bør være lang og presisere at det er teknologianalyse, slik at den blir mest mulig akademisk riktig.
- Raymond anbefalte ca. 30 sider per pers på hovedrapporten
    - Vi tenker 80-95 uten vedlegg
- Ha med kode for crawler i hovedrapport
- Struktur på hovedrapport: start med bakgrunn
- Lage et Kanban board med våre største frykter og starte med det vi synes er vanskeligst
- Skriv i hovedrapporten at vi har tatt noen forhåndsregler for å ikke komme inn på noen skikkelig fæle sider
    - Dokumenter det og ta kontakt med kripos dersom vi skulle komme over noe som er ulovlig
- «Alltid en menneskelig intensjon bak det som skjer» Raymond 23.02.23

03.03.23 - Statusmøte med oppdragsgiver

Deltakere

- Raymond Hagen (oppdragsgiver)
- Mats Dimitri Jensen
- Martin Hyldmo
- Marte Jørgensen

Agenda

- Statusoppdatering
- Spørsmål
- Annet

Statusoppdatering

- Det gruppen har gjort siden sist: Hovedsakelig kommet i gang med skriving på rapporten, bakgrunn, teknisk oppbygging og kodingen. Slik at vi ikke sparer alt til slutt. Lurt å dokumentere mye underveis.

- Vi har oppdaget at det ganske nylig har vært DDoS på dark web, så mange sider har beskyttet seg mot crawling, og at vi må gjennom ganske mange steg før vi kommer inn på selve sidene.
    - Raymond er ikke kjent med dette, men anbefaler oss å prøve å hente så mye informasjon om dette ddos-angrepet, hvem, hva, hvordan, hvorfor? Hvordan påvirker dette vår oppgave? Stagnerer progresjonen vår?
    - Noen ønsker tydeligvis ikke at folk skal ha tilgang til denne tjenesten. Russland-Ukraina krigen?

Spørsmål

- Bør vi ha med teknologier vi har valgt i teoridelen?
    - Ja, det er lurt

Annet

- To av sjefene fra Kripos stiller opp fra NC3 for å fortelle om sitt arbeid. Dato kommer senere!

- CrowdStrike: De ville ta en prat med R, kanskje vi får en liste med APT relaterte .onion sider! Evt. kanskje fra Kripos.

- Raymond og oss er interessert i hackermiljø. Kripos er kanskje mer interessert i det vi egentlig ikke vil finne? Det finner vi ut av.

- Har vi tenkt på flytdiagram for hvordan vi samler inn informasjonen? Kanskje det hadde vært en god ide. Altså en generell «mal» for hvordan vi skal strukturere arbeidet med høstingen. Anbefales.

- Spørsmål om vi har fått laget skjelett til oppgaven og det har vi sånn høvelig, men burde få på plass dette ila 14 dager så vi kan gå igjennom det med Raymond. Her snakker vi overskrifter og evt underoverskrifter i rapporten.
-
- ChatGPT hjelper til med LaTeX ifølge Raymond, hvis vi står fast.

- Tor er et nettverk som er i en hype fase, folk vet egentlig ikke hva de kan bruke det til, bortsett fra den arabiske våren har det hovedsakelig vært knyttet til ulovlig aktivitet. Det er et såkalt «våpenkappløp». Må forvente at det er land/grupper som er negative til dette og ønsker at Tor skal tas ned.

- hovedsak bare funnet hobby hackere hittil. Raymond svarer:
  "Ikke «diskriminer» hobby hackere, burde være interessert i alt som har med hackere å gjøre. For disse stedene/forumene disse holder til er et samlingssted."

10.03.23 - Statusmøte med oppdragsgiver

Deltakere

- Raymond Hagen
- Martin Hyldmo
- Mats Jensen
- Thugitha Kanavathi

Agenda

- Kort statusoppdatering
- Tilbakemelding på rapportskjelett
- Div spørsmål

Statusoppdatering

- Fortsetter skrivingen
- Har laget en spørreundersøkelse og har 190 besvarelser
- Opplever noen vanskeligheter med å crawle

Tilbakemelding på skjelettet

- 1.7 hva måtte læres: usikker på om dette bør være med, hele arbeidet er egentlig metodekapittelet.

- Metode bør være før teori. Først har du innhold og rammer, så blir det naturlig med metode, på hvordan vi har gått frem, kildebruk, hvordan gjennomføres forsøkene og hvordan/hvorfor har vi lagde crawler.
- Deretter kommer teoridelen.

- Metode-delen
  Dele opp metoden i 3 deler:
  Litteraturstudie
  Spørreundersøkelse
  Oppbygging av crawler

  Disse 3 inndelingene bør også følges under diskusjonsdelen

- Diskusjonsdelen
  Plukke ut 2-3 spørsmål som er mest relevant for problemområdet vårt og diskutere disse. Er det mulig å endre perspektivet på the dark web?

  Lage forslag til problemstilling innen neste møte:
  · Hva brukes det mørke nettet til?
  · Hvorfor finnes det mørke nettet?
  · Hva kan man gjøre for å forbedre ryktet til det mørke nettet?

- Får dere noe ut av å ha med "evaluering av gruppens arbeid"?

- Avslutning er som regel abstract, du har et arbeid du skal løse, og beskriver løsning, diskuterer fordel/ulemper, stopper der, oppsummering på alt blir avslutningen, altså samme som abstract.

- Fordeling av arbeid er bra å ha med, men evaluering av eget arbeid er sensor sin jobb. Hør litt med Erik eller veileder om hva de forventer at vi har med her.

- Kan gjerne ta med det tekniske som biblioteker og slik som blir brukt i koden i teoridelen.

- Kan sammenligne resultater fra spørreundersøkelsen og kunnskap fra litteraturstudiet. F.eks. at ut ifra resultatene på spørreundersøkelsen, så stemmer dette, men ut ifra teoridelen har vi funnet ut av dette.

- Kalle oppgaven "The Onion Routing Protocol ..."

- Hadde vært interessant å spurt om dere kunne kontrollert om det hadde vært ærlige svar.
- Spenstig å ta med en spørreundersøkelse, da folk ikke nødvendigvis svarer sant. Det er et såpass stort antall at data må evalueres ordentlig. Spørreundersøkelsen bekrefter at det mørke nettet har dårlig rykte, dette kan være grunnen til at så få har turt å prøve det. Folk flest ser ikke på dette som et normalt verktøy. Vær kritiske til om spørsmålene har vært ledende.

- Teoridelen
  Endringen fra en helt lovlig protokoll, til at det brukes til mye ulovligheter. Det mørke nettet kan brukes for å unngå sensur, mye legal bruk, men dette er mindre kjent. Prøv å få belyst dette. Dette er et verktøy som kan misbrukes. Opprinnelse; hvor gammelt er det mørke nettet?
  Hvorfor skjedde endringen? Lovlig verktøy til ulovlig plattform, hvordan skjedde denne overgangen?

    - Skriv mer om den arabiske våren i teoridelen, der det mørke nettet bruktes til å dele demonstrasjoner osv. Kommer det mørke nettet til å komme over fasen der folk er redd for å bruke det?

17.03.23 – Statusmøte med oppdragsgiver

Deltakere

- Raymond Hagen (oppdragsgiver)
- Martin Hyldmo
- Mats Dimitri Jensen
- Thugitha Kanavathi

Agenda/spørsmål

- Oppsummering av kripos (NC3) møtet
- Statusoppdatering
- Definering av problemstilling

Oppsummering av kripos-møtet

- Hva fikk vi ut av det?
    - Fikk bekreftet at det å navigere seg rundt og finne nyttig informasjon på det mørke nettet er svært tidkrevende
    - De drev mye med manuell etterforskning selv
        - bruker mye ressurser på overvåkning av det mørke nettet, blue on blue
            - Sier noe om måten det håndteres
    - Teknologien er sikker, men det er ofte menneskelige feil som avslører identiteten til brukerne

Forslag fra Raymond

- Forslag på tittel The onion routing protokoll også kalt det mørke nettet
    - En studie om the onion routing protokollen

- Fokuser på at teknologi er noe som kan brukes og noe som kan misbrukes
    - Det kalles mørke nettet av en grunn
    - Inngangsverdi er vesentlig (altså vinkling og perspektiv er viktig)
    - Det finnes mange positive sider ved det mørke nettet

- Kripos nevnte det er menneskelige feil som avslører deg og ikke teknologi, viktig å ha med i rapport

Spørsmål

- Hvordan gjør vi det med kildehenvisning når vi bruker informasjonen fra møtet med NC3?
    - Vis til uformell samtale med kripos på teams med to etterforskere. Oppgi dato, plattform og navn på etterforskerne

Statusoppdatering

Vi har klart å finne ransomware sider, Raymond sier det er kjempefint

Definere problemstilling

- Hva er egentlig det mørke nettet?
    - Null hypotese; det er noe negativt knyttet til det mørke nettet
- Spørreundersøkelse: Hvordan oppfattes det mørke nettet
- Hovedspørsmål: Hva er det mørke nettet?!
    - Bakgrunn og historie
    - Oppbygging av crawler
    - Spørreundersøkelsen: Hva er andres oppfatning av det mørke nettet

Oppdelingen som brukes i metode og diskusjonsdelen kan også brukes når vi definerer problemstilling. Vi har et hovedspørsmål og 3 spørsmål som setter søkelys på de tre aspektene av oppgaven. Dette er noe vi bør diskutere og få avklart da det kan bidra til at vi klarer å holde oss saklige gjennom hele rapporten.

31.03.23 – Statusmøte med oppdragsgiver

Deltakere

- Raymond Hagen (oppdragsgiver)
- Mats Dimitri Jensen
- Thugitha Kanavathi
- Marte Jørgensen

Agenda/spørsmål

- Føler vi står litt fast: Har i løpet av tiden fått en del kontrabeskjeder, fokuser på dette og neste gang dette. Mistet litt oversikt/perspektiv. Vi trenger å få på plass en problemstilling. Må vi ha forskningsspørsmål/hypotese?? - spørre Erjon
- Teori: Hva kan vi ha med der? Teori på Dark Web, tidligere forskning? Holder bakgrunn/historie/teknisk?

- **Lage noen forskningsspørsmål, hva er det vi prøver å besvare? 3 ting:**

- Generelle bakgrunn, hva er protokollen, onion, hva er deep web. Teoretisk, sette standarden.. R har et ønske om at vi skal motbevise en del myter, ikke alt er så ille på det mørke nettet.
- Utviklingen av programmet - "enkelt" - dokumentere i oppgaven utviklingen av et program er. Sekvensdiagram, skisser, tanker osv., møtt på utfordringer, dokumenter dette. (Hva kreves for å hente ut data fra det mørke nettet/protokollen
- Uformell/formell analyse av det mørke nettet og teknologien. Ikke nå galt i å ha en nullhypotese som sier folk oppfatter at det mørke nettet er en negativ teknologi? Gjøre lett analyse av innsamlet data, lete i databasen om det finnes "prejudice about dark web" lignende ting. Få frem "det folk ikke forstår er de redd". Der vi har hentet info og har erfaring er det å skjønne hvorfor folk er redd.

- Vi har en del info, så vi kan stille spørsmål vi vet vi kan svare på.
- Lag noen underspørsmål som kan belyse mer – utdype mer

Forteller en historie om hjernen og impuls, reaksjon, følelse - oppfattelse.
De som ikke kjenner bakgrunn og ikke har satt seg inn i teknologien vil kanskje ha fordommer, men vi som nå kjenner det så skjønner vi at det ikke bare skjer kriminelle ting der. Det er en legitim bruk, men vanskelig å skjønne for de som får servert en svindlerhistorie der hackerne har solgt info på dark web.

- Årsaken til at vi skal gjøre denne oppgaven: Raymond tror det finnes indikatorer man kan finne for å lete etter ting som kan identifisere avanserte trusselaktører.

- Knytte opp spørreundersøkelsen til det tekniske (Erjon har uttrykt dette)?
  Den åpenbare koblingen mellom dette er at alt starter med et menneske. Hovedresultatet av spørreundersøkelsen er at folk er skeptiske til det mørke nettet. Dette handler om "fantasien" til enkeltindivider. Og så er det om disse har kunnskap eller ikke. Dette er påstander som er med å påvirke bruken? Bevist at teknologien har et tvilsomt rykte, men det er mennesker som har fordommer mot de menneskene som bruker teknologien.

I likhet med andre teknologier som kommer som har hatt et dårlig rykte, så er det veldig vanskelig å endre til at det blir noe positivt. Det blir dratt ned og vanskelig å endre rykte.

- Må dataen analyseres og sammenlignes med annen forskning?

  Gjøre en sammenlignende analyse om noe som er antatt å være negativ? (Bitcoin). Teknologien er besudlet, folk tror det hovedsakelig brukes til svindel?
  *Eventuelt noe annet de har en mening om, men som de ikke egentlig forstår.

Quote "Intelligence is the ability to adapt to change" av Raymond.
I hovedsak – tenk på disse spørsmålene nevnt over. Se på disse og prøv å svar på disse for å få mer kjøtt på beina.

14.04.23 – Statusmøte med oppdragsgiver

Deltakere

- Raymond Hagen (oppdragsgiver)
- Martin Hyldmo
- Mats Dimitri Jensen
- Thugitha Kanavathi

Agenda/spørsmål

- Skal vi skrive at det er litteraturstudie?
    - Vi har ikke søkt like systematisk som vi burde ha gjort?
- Hvor bør tjenesteangrep avsnittet flyttes til?
- Bør teori-delen inneholde noe mer?

Starter med gjennomgang av hovedrapport der vi går gjennom tilbakemeldinger.

Oppgavebeskrivelse: I tillegg til faktiske oppgavebeskrivelse kan dere fortelle "vi ble enig med oppdragsgiver å gjøre slik og slik ..."

Læringsmål: Grunn til at dere er gjennom alle forskjellige typer studier er fordi dere ønsker å få en utdypende kunnskap innenfor TOR

Målgruppe: Digdir kan ha en interesse av å bruke rapporten som et opplæringsverktøy for TOR

Bakgrunn: Beskrive at dere har en bakgrunn som kan løse dette problemet. Deres utdannelse gjør det mulig å fullføre oppgaven på en trygg måte. Greit å få med – sikkerhet er mye mer enn teknologi, dra nytte av det dere har lært.

Hva måtte læres: denne kan være skummel, svakt om man ikke kunne noe fra før og om man kunne alt fra før. Programmeringsmessig vil det være bedre å forklare at man må kunne TOR teknologien for å lage scraper i stedet for at man måtte lære Python.

Dere valgte Linux lite fordi jeg anbefalte det, og fordi de pc-ene dere fikk var relativt svake, unnlot å laste ned visse pakker, de gjør jobben innenfor å kunne koble til TOR og lage/kjøre scraper. Dere må ha en viss forståelse for hvorfor dere tok alle valg, og forklare dette.

Teori: finne ny tittel for dette? Knytte selve oppgaven opp mot denne tittelen – tittel må være beskrivende, f.eks. Teorien bak TOR, Teoretiske momentet for oppgaven.

Bakgrunn og historie, bra innhold, men så tidlig som mulig forklare hva dere mener med åpne, dype og mørke nettet. Definer så raskt som mulig. Splitt opp slik at det blir naturlig overgang. Definisjon skal gjerne gå før bakgrunn og historie. Starten av 2.2 kan kanskje puttes inn i 2.1.

Dere har en del referanser, men virker som at det er litt tilfeldig hvor dere bruker referanser (?), finn en balanse mellom å nevne alle referanser eller ta det litt mer rolig. 1 kilde er plagiat, 10 er forskning.

Den arabiske våren kom veldig plutselig. Kanskje dere vil ha en egen om bakgrunn, og ha en egen del med historie?
Praktiske eksempler knyttet til TOR kan være et eget kapittel, med den arabiske våren som kommer inn der. Den Arabiske våren er ikke eneste grunn til å velge TOR, men er et godt eksempel på fine sider med dette.

Teknisk oppbygging: denne tittelen forklarer ikke hva slags teknisk oppbygging dette er.

Bildet om kryptering er dessert, pass på at det ikke kommer så tidlig.

Tjenestenekt: kan definitivt ha det med, men når dere er i metode, så observerte dere et tjenestenektangrep, dette sier noe om trusselbildet til TOR, det at du kan bruke torify eller noe til å kjøre TNA .... det er et såpass komplekst tema, kanskje ha et eget kapittel der dere har beskrivelser av definisjoner. Etter kapittel 1. kanskje at 1.11 er definisjoner.

Søkerobot: Den fortjener et eget kapittel, det er en viktig del av oppdraget. Lag et eget kapittel 3 som heter søkerbot, med innledning, bakgrunn, hva har dere lagd, biblioteker, begrunnelser. Dere viser at dere ikke trengte å lære dere å programmere, men at dere trengte å lære dere å programmere opp mot TOR. <- slike ting gjør at dere kan komme opp mot en A

Skrive metode kapittel- i denne oppgaven prøver vi å løse 3 problemer, litteratur, søkerobot og spørreundersøkelse, beskriv litt rundt metoden, men for detaljer, se kapittel om søkerobot. Henvis til detaljer i deres eget kapittel.

Datainnsamling: vi har brukt 3 forskjellige metoder for å løse oppgaven, fortell hvorfor, detaljer på hvordan dette ble gjort finner dere i ref{...} og dette ref {…} osv.
Trenger ikke følge IMRAN 100%, men å bruke et enkelt oppsett er alltid bra. Selv om noe har virket helt ubrukelig, ha det med i referanselisten, det viser at dere har prøvd.

Undersøkelse: Se litt på hvordan dere vil gjøre det med spørreundersøkelsen, kanskje bedre å putte bilder som vedlegg. Aldri ha en henvisning til en figur uten å ha en begrunnelse. Tanker, gjerne hypotese, hvor mange som har spurt, hvor mange svar osv.

For å rettferdiggjøre 3.1 hva er årsaken til at vi valgte akkurat disse spørsmålene, gjøre med alle spørsmål, og observasjoner og konklusjoner på slutten av kapittel om undersøkelse. Vær varsom når dere bruker disse svarene, kan alltid være noen som ikke svarer sant.

Observasjoner skal i resultat. Kakediagram og slikt fra undersøkelse.

Søkeroboten: gi en liten teaser tilbake på at dette var oppgave, og at oppdragsgiver ønsker data fra det mørke nett for videre forskning.

Trends bilde: Begrunn med at dere kan disse programmeringsspråkene fra google trends bildet.

Før oppbygging av søkerobot ha forklaring på alt av oppbygning rundt, VPN, maskiner, linuxlite.

Oppbygging av søkerobot: Her må det inn å beskrives hva er selenium og andre bibliotek.

Etter flytdiagram kan dere starte med datainnsamlingskapittel. Lag et flytdiagram på etterforskningsfasen, hvordan dere finner sider som dere skal scrape. Beskriv utfordringer, DDoS, visse sider blir tatt ned eller har stoppet å fungere, de har CAPTCHA, vis at dere har brukt tid på å forstå alt av slike ting.

Resultater: Her er der bare å tenke interdisiplinært. Vis frem kunnskapen deres, dere skal kunne forklare noen hva TOR er. Spør Erjon om resultater delen. Resultater er litt det samme som Abstract. Er det noe som er unaturlig å ha med i metoder og slik, så hører det hjemme i resultater. Enten må Resultater være såpass lagt at det får inn noe nytt, ellers blir det bare en abstract. Er det noe som ikke er fortalt, så må det inn i resultater.

Oppsummering skal gi deg konklusjonen, og resten skal oppbygge denne konklusjonen. Diskusjonsdelen skal peke tilbake til spørsmål man har tidligere. Er TOR noe som bare kan brukes til illegale ting, eller kan det brukes til noe legitime ting? Slike typer diskusjoner. En observasjon vi ser i spørreundersøkelsen, er at folk som aldri har brukt TOR fortsatt mener at det er farlig.

Kan være lurt å lage en inndeling med en gang, og skrive etter den. Sjekk om Erjon synes dette ser bra ut. Ta en test om dere har nok informasjon.

21.04.23 – Statusmøte med oppdragsgiver

Deltakere

- Raymond (Oppdragsgiver)
- Alle gruppemedlemmer

Agenda

- Statusoppdatering
    - Fått mer føringer for struktur
- Annet

Statusoppdatering

Gruppen oppdaterte oppdragsgiver på hva som har blitt gjort og hvordan uken har vært. Vi har fått gode innspill fra veileder på hvordan oppgaven kan utformes og tror dette kan løse problemet med sammenheng i oppgaven. Veileder har også kommet med forslag til plasseringen av tjenestenektangrep-avsnittet, som ble luftet med oppdragsgiver også.

Annet
Oppdragsgiver spurte etter oppdatering om data som vi har klart å høste. Det har blitt funnet noen vulkan filer (som ble nevnt av oppdragsgiver forrige møte) og en zip-et versjon sendes over til oppdragsgiver på e-post.

28.04.23 – Statusmøte med oppdragsgiver

Deltakere

- Raymond (Oppdragsgiver)
- Martin Hyldmo
- Mats Jensen
- Thugitha Kanavathi

Agenda

- Statusoppdatering
- Forslag/tips fra oppdragsgivere
-

Statusoppdatering

- Fortalte litt om hvordan vi ligger an og hva vi har gjort denne uka

Forslag/tips

- Tipser om at vi skal bruke chatgpt til å strukturere tester

- Anbefaler å bruke chatgpt og lime inn et ferdigskrevet avsnitt og be chat om å formulere avsnittet slik at det passer inn en vitenskapelig skrivemåte, den skal være i nøytral form og fokus skal være på god flyt og lesbarhet.

- Sjekk redaktør funksjonen på Word til å sjekke plagiering

- Sender over 2.utkast til Raymond i løpet av neste uke

- Bruke data som vi har høstet i resultat delen i rapporten

- Søkerobot delen: to komponenter
  - Produktutviklings fase
  - Etterforskning: hvor mye data er tilgjengelig for alle?
    - Det er ikke alt som er like tilgjengelig som folk tror
    - Mekanismen er undervurdert. Det krever ganske mye å få ut data fra det mørke nettet

- Forskningsspørsmål: uthenting av data

  - Bruk av scraperen er krevende, å komme seg inn og få tilgang til data
  - Kan sees på som et eksperiment? Skrive at vi har prøvd og feilet …. resultatet er …...

- Nytteverdi av Tor?

Spørsmål

- Brukes det en eller flere mellomnoder i Tor-nettverket?
  Raymond skulle undersøke, men Mats fant dette: https://support.torproject.org/misc/misc-11/

05.05.23 – Møte med oppdragsgiver

Deltakere

- Raymond Hagen
- Mats Jensen
- Martin Hyldmo
- Marte Jørgensen

Agenda

Gå gjennom kommentarer oppdragsgiver har til 2. utkast.

Referat:

Bør vi skrive i introduksjon at det blir brukt en del IT-terminologi, og dersom man ikke har forkunnskaper innen IT, så bør man sjekke ordlisten før man starter?
Ja høres fornuftig ut ifølge Raymond.

Vi har litt jobb igjen ifølge Raymond.

INNHOLDSFORTEGNELSE: Et vedlegg som heter "Flytdiagram", fint det. Men burde være litt mer beskrivende, hva slags flytdiagram.

Et spørsmål fra forrige gang: Bruke det flere enn en mellomnode? Det vet vi ikke, fordi det er det som er løkrutingen.

Overskrifter må på plass!
Vi må bruke mye tid på å lese igjennom, setningsoppbygging, ord, uttrykk osv. Burde bruke mer enn 1 dag på dette.

Kapittel 1 små avsnitt om skrivestil, er det noe som kreves eller har vi valgt selv? Det var et forslag på et lynkurs.

Teori: Introduksjon til teori: Mer informasjon om hva dette kapitlet skal fortelle oss. Det som stod der nå sa ikke Raymond så mye mente han.

Figur 2.1 - kilde, var et mellomrom som hadde sneket seg inn. Funker nå. Se litt på figurer og kilder, men normalt sett bruker det å stå Foto: sted hentet fra.

God jobb i form av litteraturstudie, hente inn informasjon. Innholdet er veldig bra, men det som kanskje er enn største utfordring er at vi har veldig stor variasjon. Vi har litteraturstudie, spørreundersøkelse og innhenting. Dette skal gi et resultat. Da må rapporten bære preg av at vi har gjort forskjellige ting. Metodemessig er det bra. Vi har gått bredt ut. Men vi må hvor skal vi med disse tre metodene, hva vil vi frem til. Hva har gjort at vi kan nå si at vi kan mye mer om Tor enn vi kunne før.

Slå sammen bakgrunn og historie igjen, R hadde glemt at han har bedt oss ta det hver for seg.

Tjenestenekt – hvor skal det legges? Ok nå. Det vi beskriver i 2.2.3 er på en måte en erfaring vi har gjort gjennom prosjektet, ikke en generell beskrivelse av Tors gang historisk sett! Det må påpekes.

2.3 Veldig fint kapittel, bra innhold. Tenk igjennom overskrifter. *Utforsking fordeler aspekter* -->
Dobbelt opp. Forslag: "Er det bare tvilsom aktivitet Tor er brukbar til?" ELLER "Dette kapitlet utforsker det vil allerede vet, årsaker til at Tor er en fornuftig løsning"

2.3.1.1 er det egentlig et eksempel om ytringsfrihet. Arabiske våren.
Samme med Russland, er det egentlig enda et eksempel?
Er det egentlig bare et underkapittel på noe som går på sensur? Skjønte ikke helt hva han mente?

2.3.2 Uavhengighet og frivillighet – tittelbytte. Klar og tydelige titler generelt, som sier nøyaktig hva det er som kommer under dette kapitlet. Legge oss på nivå med 2.3.3 i tittellaging.

2.4 teknisk oppbygging (av Tor nettverk – ny tittel). Burde dette komme tidligere? Samtidig, alt kan jo heller ikke komme først.
Beskriver hva det egentlig vi prøver å si. Gjøre et valg i forhold til oppbygging.

- Veldig mye tekst i 2.4. Hadde det hjulpet å finne en referanseskisse fra Tor Prosjektet for å vise hvordan Tor fungerer? Slik at vi kan referere til en figur mens vi skriver.
- Har mange moduser, hvorfor har vi kun med https-mode-only. Er det tilfeldig eller tenker vi at den er så grunnleggende at den er valgt som eksempler. Skrive om andre moduser kort i et annet avsnitt.
- Onion domener: En figur som forklarer (samme figur som er nevnt over). En tegning som viser at hvis du skriver xxxxxx.onion vil du ikke komme inn på den fra vanlig webleser, men kun via Tor. .. Bryt opp teksten litt for å ha forklarende figurer.
- Kryptering: Er fint at den står for seg selv. Den er grei, men det er heavy tegning som må beskrives litt med omhu.
- Figur 2.3 gir en grei oversikt som er beskrivende enn 2.2 - burde bytte plass/komme før.

2.5 Søkerobot. Bør dette kanskje være et eget kapittel. Raymond spør oss, men virker som han er litt for det.

- Tidsbruk. Hvor mye tid har vi brukt på koding (kanskje 20-30%). Såpass stor del at det må vises! Arkitekturen, hvorfor vi har laget den, hva har blitt gjort, hvordan – det er viktig at kommer frem. Må komme frem at det er brukt mye tid og jobbet mye. Bibliotekene har vi lært på grunn av Tor er laget som det er og at det måtte læres slik at søkeroboten kunne tilpasses Tor.
- Flytter det på riktig sted.

3 Metode. Han mener vi burde bytte plass på teori og metode. Kanskje vi skal tone ned teorien, til kun bakgrunn og historie og noe rundt det. Slik at leser får innblikk i hvorfor vi gjør denne oppgaven og kunnskap til å skjønne det viktigste. Leseren har lest 20 sider og så først kommer metoden om hvordan det er gjort.

- Tabell litteraturfunn: Referere til tabellen i teksten når vi forklarer!!
- Åpne, dype, mørke nett – skrive noe om fremgangsmåte. Igjen, burde kanskje tone ned i teorien og så komme inn på detaljene i metode. Sitter med inntrykk av at det samme han har lest før kommer igjen, historie, teknisk oppbygging osv. blir repetert.
- 3.3 er bra. Viktigste å få frem at hvorfor og hvordan det er brukt
- 3.3.1 vedlegg istedenfor i teksten mener Raymond.
- 3.4 Veldig grei. Det eneste Raymond savner er det mest åpenbarte. Etter kriteriene hva kan vi? Det er mange ting vi måtte klare selv om vi kunne programmeringsspråket. Skrive at vi kunne flere språk, men valgte Python. Selvskryt!
- 3.4. traversering - Bra, men. KRYPEPROSESS??? Nei, kall det noe annet.
- 3.4.3 - 3.4.4: Rekkefølge på avsnitt. Tor er brukt til to ting. På grunn av oppdragsgiver ikke ville at vi skulle oppleve anonymitetsbrudd fikk vi 4 PCer og i tillegg brukte VPN (betalt selv), ha en tydelig distanse, slik at vi ikke skulle være redd for angrep.
- Figur 3.5 - navnet er litt forvirrende under?

4. Resultater: Bare for lesbarhetens skyld. Sensor leser introen og så midten (der mister de fokus), så resultat og vurdering (da tar de seg sammen igjen). Viktig å skrive godt her.

- 4.1 Husk å repeter spørsmålene vi skal svare på, slik at leseren slipper å gå tilbake til intro.
- Hva er våre tanker rundt Tor nettverket. Bra det som står under 4.2 spørreundersøkelse. Hvordan syr vi sammen resultatene av spørreundersøkelsen med litteraturstudiet og utviklingen? Hva er et klart resultat vi har fått: At flesteparten oppfatter det mørke nettet som et skummelt område --> Litteratur: skummelt blant befolkning, men finnes mange gode formål --> Innhente data = ? Hvordan kan dette knyttes sammen, disse resultatene.
- Det krever mange ressurser og kapasitet og tilpasning for at vi skal klare å hente ut data. Kripos-møte - bruke eksempler, at man ser hvor mye tid og ressurser de bruker – ikke rart at vi har brukt så mye tid og likevel ikke oppnådd et fullstendig resultat - må komme frem i diskusjonen!
- Utfordring i form av at teknologien blir oppfattet som et problem, negative assosiasjoner. Det finnes eksempler på at i de rette hender så er mekanismen en funksjon.

Hovedfunnet vårt er: oppfatningen, hva er det egentlig og hvor vanskelig er det egentlig å bruke?

Intensjon og kapasitet: Noen tilfeldigvis gikk på Tor og så ble det lastet ned barneporno, det går ikke. Viktig å påpeke, det fungerer ikke på den måten. Mannen i gata tror kanskje på dette, men nå vet vi at det ikke er slik. Her må man bevisst lete etter det man ønsker å finne.
Spørreundersøkelsen: vet ikke om de lyger, men for eksempel den som sa han hadde kjøpt narko, han vet i alle fall at det er mulig å gjøre der.

Erfaringer:
Det at vi heller skulle skrevet oppgaven på engelsk er et funn. Mange ord og uttrykk som er på engelsk som preger dette problemområde og denne verdenen. Viktig vurdering gjort i etterkant/underveis.

Videre arbeid: Raymond skal bruke det i sin doktorgrad, går mer på at han skal bruke det i sitt forskningsarbeid i forhold hvordan avanserte trusselaktører opererer. Det kan også hende at den teoretiske delen av rapporten kan brukes i Digdir for å lære folk hva Tor er.

Generelt: Tenk på navn på overskrifter. De skal skjønne hva som kommer i det de leser overskriften. Oppbygningen og plassering av ting. Gir det mening for leseren, kommer det i en logisk rekkefølge. Viktig med god flyt.

12.05.23 – Møte med oppdragsgiver

Deltakere

- Raymond Hagen
- Mats Jensen
- Martin Hyldmo
- Thugitha Kanavathi

Agenda

- Statusoppdatering
- Oppgaven vi fikk stemmer ikke lengre med det oppgaven har blitt til. Oppdragsbeskrivelsen nevner hovedsakelig statsfinansierte hackere og ønsket data om disse med tanke på forskningen din (derav bakgrunn og problemområde). Oppgaven vår er nå blitt til hva er Tor, hensikt, oppfattelse i befolkningen og teknisk oppbygging + utforskning av fordeler med Tor (som er hele litteraturstudiet vårt). I tillegg høsting av data om ulovlig aktivitet, hovedsakelig ransomware. Vi får ikke dette helt til å henge på greip, hva gjør vi?

Det viktigste for dere er å formidle at dere har lært noe, dere har fått rammer om at det er mørke nettet dere skal utforske, og jeg ville gjerne ha litt data. Utfordringen her er en vag oppgavebeskrivelse, og en del utfordringer gjør at oppgaven vil forandre seg underveis.

Har hatt god dialog med R hele veien, så bare å være ærlig med at oppgaven var «vanskelig, vag, lite spisset», dette er en omfangsvurdering, og ut ifra det kan dere vise til hvorfor det er avvik fra det dere starter med, til det dere endte opp med.

Vi har gjort alt som skulle gjøres, men vinkling har snudd en god del, gått fra hackergrupper og ransomware-grupper og om til mer positive sider ved Tor.

Vet ikke hva jeg skal foreslå, men bare prøv i oppsummering at oppgaven har blitt fordreid grunnet kilder og det vi har gjort, det oppdragsgiver ville ha, har vi løst, men som en del av læringen har vi oppdaget: «vi som andre hadde litt negative tanker rundt Tor, mye skummelt, men etter vi har satt oss inn i det, vet vi at det kan brukes til mye godt, og at det er ikke bare skumle saker der.» Og av denne grunn har vi endret oppgaven litt over til at Tor kan brukes til både godt og vondt.

Dere har svart på spørsmål, men sjekk om dere har et spørsmål å henge dere på.
Tips til forskningsspørsmål: «dere var negative, men ser at det kan brukes på godt og vondt», spørreundersøkelsen viser til at flere tenkte slik som dere gjorde i starten.

Oppgaven vil derfor holde et nøytralt dypdykk i løsningen (ikke si at det er positive og negative)
Fjerne derfor vil særlig positive sider belyses. <- ikke objektivt.

For å få til dette få en forståelse av TOR, hva Tor protokollen kan benyttes til.

Hele tiden er hovedpoenget at dere skal lære mest mulig om Tor. (som mine veiledere sier til meg «det er bare å skrive det»)

Dere har gått fra å være negative mot Tor til å bli mer nøytrale. Det er nøytrale dere skal være.

Hadde dere lagd en scraper for det vanlige nettet hadde kunnet laste ned en enkel kode på nett, og skaffet 5TB enkelt. Men dere skal gjøre dette på det mørke nettet, og her er det en del utfordringer. Hvorfor? Jo dere må inn i protokollen å se, det finnes kanskje mottiltak osv.

Det dere faktisk skulle var det å få en dyp forståelse for Tor. Det har dere gjort gjennom alle disse fremgangsmåtene.

Ikke legg til noe nytt, dere har mye stoff, mye bra arbeid, dette blir bra! Tenk helhet, dere har den røde tråden. Tenk over egne påstander og meninger som ikke er nøytrale, de må enten forklares eller fjernes.

19.05.23 – Møte med Oppdragsgiver

Deltakere

- Raymond Hagen
- Marte Jørgensen
- Mats Jensen
- Thugitha Kanavathi

Agenda

- Siste utkast

Siste utkast

Lurt å gjøre seg ferdig og så finpusse etterpå. Ikke noe mer nytt nå. Bare fullføre det vi har. Abstract kommer helt til slutt, kortere versjon av oppsummeringen.

Mye tekst. Men det e bra. Det er gjort mye god jobb! Vi har gjort mye i forhold til struktur, det er bra det er på plass.

Ser det bra ut? De to viktigste kapitlene er abstract og konklusjon. Metodekapitlet ser de på og referanser. Det vi burde tenke på er at det er 3 ting vi blir bedømt på: Omfang, vi har gjort mye og tre veldig forskjellige ting. Her er vi innenfor. Det andre er fremstilling, at vi klarer å strukturere et fint dokument. Ser veldig fint ut nå, ryddig og bra. Pass på at vi har greie referanser og lese nøye at vi har logisk oppbygging. (Eksempel, traversering og oppsett av pc, bytte plass). Tommelfingerregel, det som krever mest for leseren er gjerne det som komme til slutt. Der vi kanskje vil score høyest er det siste. Vi har vist at vi har omsatt det vi har lært i noe nytt. På grunn av Tor er bygd som det er bygd så kan vi ikke programmere søkeroboten som vi ville gjort hvis det var det vanlige nettet vi skulle crawlet. Omsatt kunnskap til vanlig forhold. Skryt av oss selv, viktig!

Form og farge, på konklusjonen. Vi har skrevet mye og da er det å bestemme hva som skal med i konklusjonen. Hva har vi gjort, hva har vi lært, viktige funn. En ting R tenker er viktig er det at mismatch mellom hva teknologien er og oppfatningen av Tor. Et annet funn, teknisk sett har det høy brukerterskel, fordi det ikke er intuitivt å finne ting. NC3 etterforskning, bruker mye tid på det åpne nettet for å finne frem på Tor senere. Det er vanskelig å bruke en crawler på det mørke nettet.

Litteraturstudie: funn i forhold til positiv bruk. Litt viktige funn. Kan også puttes i konklusjon.

Abstract: Vi har gjort xxxxx kort fortalt. Være en teaser. En eller to av de viktigste funnene nevnes, og så gir vi et hint til hva rapporten skal gjøre.

Erfaringskapittel: Prøv å vær litt mer positiv! Det skal være utfordrende. Veldig viktig i erfaring - språk. Flytt opp, mer viktig. Tenk litt over det: Ikke snakk ned oss selv, vi har jobbet bra. Vi har svart ut de spørsmålene Raymond hadde. Ikke gi inntrykk av at vi ikke fikk svart ut noe.
Tre metoder, spesiell oppgave, problematisk: skriv om. Siden oppdragsgiver hadde klare mål om hva han ønsker tilbake. Siden teknologien var ukjent for oss, så var det flere problemstillinger og metoder som måtte forsøkes før vi fant den som passet/fungerte best.
Gantt-skjema ikke fungerer, er som forventet.
Faste møter med Raymond, med å sørge for fremdrift hele tiden, selv om tidskjema blir endret. Vi er også interessert i å vite om vi har gjort det jevnt eller ikke. Kanskje ha med en setning eller to om nettopp det at vi hadde en utfordring med at ting vi trodde var enklere tok lengre tid, fordi vi måtte gjøre en del utforskning underveis.

Kult at vi tok oppgaven på strak arm, selv om det har vært mye nytt. Oppgaven er nyttig for Raymond. For eksempel det at man bruker tid på det mørke nettet, men at det faktisk ikke bare er fæle ting der. Mer nyansert bilde enn man tror.
Vi har fått en god del erfaring som Raymond tenker er nyttig i forhold til at teknologien som ligger i bunn er mer kompleks enn man tror.

Fokuser på rekkefølge, får med oss leseren underveis. At det virker logisk.
Innholdet er bra. Språkvask må til.
Få skrevet konklusjon.
Erfaring, fokuser på det positive. Det er en vanskelig oppgave, men vi har lært mye. Det er en oppgave vi var ambivalent til det, fordi vi den var i det ukjente, og at den ble litt annerledes enn det som var tenkt, men gikk bra til slutt.
Skyld på Raymond hvis det er noe vi mener er negativt. At det var utenfor vår kontroll for eksempel.

Overskrift: xxxxxxx Tor (The onion router/routing protocol)

Vi ha gjort mer enn Raymond hadde trodd. 😊

Vedlegg K

Møtereferater fra møter med veileder

Møtereferatene fra møter med veileder samt noen interne møter.

11.01.23 – Seminar: Lynkurs i prosjektstyring

Deltakere

- Tom Røise (foredragsholder)
- Alle gruppemedlemmer

Mål

Komme i gang med bacheloroppgaven og lage plan for gjennomføring

Agenda

- Karakteristika ved Bacheloroppgavearbeidet
- Prosjektplanen – gjennomgang av mal
- Valg av prosessrammeverk / systemutviklingsmodell
- Råd og tips til prosjektstyring i Bacheloroppgaven

Oppgaver

Få ferdigstilt prosjektplan innen 31.janua

20.01.23 – Statusmøte med veileder

Deltakere

- Erjon Zoto (veileder)
- Marte Jørgensen
- Martin Hyldmo
- Mats Dimitri Jensen

Agenda/Spørsmål

- Skal vi kontakte politiet i tilfelle vi laster ned ulovlig innhold?
- Prosjektrammeplan, hvilken kan passe oss best?
- Nye møter, faste tider?
- Noen ideer til VPN?
- Når vi skal kjøre crawler, er det mulig å bruke skolenettet til dark web?
- Problemstilling på oppgaven vår?

Politi

- Når det gjelder å kontakte politiet før vi skal starte eksperimenteringen foreslår Erjon at vi spør Raymond, da han ikke er sikker hva som er lurt å gjøre.

Prosessrammeverk

- Angående prosessrammeverk for oppgaven fikk vi ingen konkrete forslag, men fikk beskjed om å se på hva Tom Røise hadde lagt ut. Siden vår oppgave er en blanding mellom research og utvikling er det ikke like lett å velge prosessrammeverk. Må kanskje velge en blanding og tilpasse den. Vi må se mer på dette selv.

Veiledningsmøter

- Videre veiledningsmøter blir torsdager kl. 10 Vi booker rom i forkant (Virket som Topas var ønskelig for Erjon). Legger ut eventuelle relevante dokumenter i Teams gruppen i forkant av veiledningsmøter.

VPN

- Erjon visste ikke annet enn det som NTNU har for alle studenter, så ba oss høre med Raymond.

Skolenett

- Dark web på skolenett: Usikkert, spørre noen andre. Vi tenker at vi spør IT avdelingen på NTNU.

Problemstilling

- Når det kommer til problemstilling så sa Erjon at vi kanskje burde ha noe med å hente ut data, utvikle til å gjøre det automatisk?

26.01.23 – Statusmøte med veileder

Deltakere

- Erjon Zoto (veileder)
- Marte Jørgensen
- Mats Dimitri Jensen
- Thugitha Kanavathi

Agenda

- Tilbakemelding på prosjektplan
- Sluttrapport
- Rammer

Tilbakemelding på prosjektplan

- Understreke at vi skal forstå dark web og at høste data er nummer 2 på lista. Forholde oss til dette bevisst hele veien.
    - **Problemområde**
        - Skrive noe mer om APT og definere det. Bruk kapittel 8 i digital sikkerhet boka fra førsteklasse for å lese om APT
    - **Rammer**
        - Utdype mer, skrive mer under hver undertittel. Skrive litt mer detaljert under tidsfrister forklare gant-skjema, frister for eksempel når vi stopper å innhente data.
        - Anonymitet, hvor dypt kan vi gå osv. Hvorfor er dette viktig og hvordan kan dette hjelpe oss med å høste data osv.
        - Ha med det at vi har fokus på statsfinansiert
    - **Problemavgrensninger**
        - Begrensinger i forhold til hvilken versjon av Tor i bruker og at vi bare tar hensyn til det vi har tilgang til nå og den Tor-versjonen vi har tilgang til nå.
    - **Scrumban (4.1.1)**
        - Legge til bilde av flytdiagrammet som beskriver prosessen.
    - **GANT**
        - Skrive noe om hver fase og «under fase» i korte avsnitt under gant-skjemaet
        - 3.1 og 3.2 bytter plass (informasjonsinnhenting før oppsett og språk)
        - Kjøre crawler allerede 20.mars (blir en del av testing prosessen, teste 10. og deretter kjøre for å sjekke om det fungerer som det skal)
        - Fase 2: Bakgrunn og historie først også 2.1, 2.2 og 2.3

Hvordan skal crawleren utvikles?

Utviklingsmetode i forhold til crawler

- Waterfall metode?

Sluttrapport

- Når vi skriver rapporten kan vi argumentere for at vi brukte crawler og hvordan den er bygd og hvorfor den ble bygd sånn som den ble.

- Jo mer refleksjon rundt avgjørelser desto bedre. Argumenter for valgene vi har tatt.

- Han var villig til å lese gjennom utkast av rapport:
    - fordele kapitler holde han oppdatert på hvilke kapitler som har blitt endret på
    - Bør være et skille mellom litteratur og utvikling
    - Sende førsteutkast – påskeferie 6./7. april
        - 1 eller 2 ganger etter det også – finner ut av det etter hvert

Rammer

Høre med Raymond(oppdragsgiver) om data som høstes hovedsakelig skal være på engelsk, eller er ikke språket så viktig? Mulig å sette opp noe filter for å oversette teksten/innholdet på siden før nedlastning/underveis/før lagring?

Oppgaver

Gjøre endringer basert på tilbakemelding og levere prosjektplanen innen 31.januar

30. 01.23 - Statusmøte internt

Deltakere

- Mats Jensen
- Martin Hyldmo
- Thugitha Kanavathi

Agenda

- Prosjektplan
- Neste fase

Prosjektplan

Avdekke mangler

- Rammer og avgrensinger har vært litt overlappende og dette har skapt litt usikkerhet på hva som skal under hva. Vi har fått dette avklart og har skrevet om litt basert på det innputtet vi fikk fra oppdragsgiver

Ferdigstilling

- Alle gruppemedlemmene som har vært til stede har korrekturlest dokumentet og har endret på diverse setninger og rettet på tegnsetting.
- Ferdigstilt versjon av prosjektplanen ble sendt inn til veileder på epost.

Neste fase

- Vi har fått avklart hva vi skal gjøre i ukene fremover. Utvikling av crawleren skal skje parallelt med informasjonsinnhenting.

02.02.23 – Statusmøte med veileder

Deltakere

- Marte Jørgensen
- Martin Hyldmo
- Mats Dimitri Jensen
- Thugitha Kanavathi

Agenda

- Oppgavefordeling

Da veileder ikke dukket opp ble møtet omgjort til et internt statusmøte.

Arbeidsfordeling

- Arbeidet med informasjonsinnhenting har blitt fordelt blant gruppens medlemmer. Vi har delt oss i hvilken form for informasjon som skal finnes. To av medlemmene skal innhente informasjon om bakgrunn/historien til det mørke nettet og teknisk oppbygging. De to resterende medlemmene skal innhente informasjon rundt crawler'en som skal utvikles.

23.02.23 – Statusmøte med veileder

Deltakere

- Erjon Zoto (veileder)
- Martin Hyldmo
- Mats Dimitri Jensen
- Thugitha Kanavathi

Agenda

- Statusoppdatering
- Råd og veiledning

Statusoppdatering

- Crawleren er 70% ferdig, testet den på det mørke nettet
- Offisielt ferdig med research fasen
- Planen er å skrive på hovedrapporten neste uke

Råd og Veiledning

- Skrive noe om hvorfor crawleren er den beste løsningen i hovedrapporten, bakgrunn for valg
- Sammenligne innholdet som scrapes og lagres i database og det som faktisk dukker opp på nettsiden for å sjekke om det er noe crawleren ikke får med seg
- Drøfte/diskutere innholdet på det mørke nettet og det åpne nettet, hva som er forskjellen, positive og negative sider
- Kan det være en idé å lagre en header-fil med informasjon om forfatteren og dato for hver nettside i databasen?
- Skal det være mulig å finne artikler og innhold ved å søke på dato i databasen?
- Vurdere å se gjennom innholdet i databasen, se på statistikk, nøkkelord på hva som er mest brukt osv.
    - Kanskje lagre brukernavn på noen brukere?
- Vær så detaljert som mulig

30.03.23 – Raskt statusmøte med veileder

Deltakere

- Erjon Zoto
- Martin Hyldmo
- Mats Jensen
- Thugitha Kanavathi

Agenda

- Statusoppdatering

Statusoppdatering

- Følge opp spørreundersøkelsen, og sammenligne denne med andre undersøkelser.

- Hvordan ligger dere an med søkerobot? - Måtte starte fra scratch, og bygge en helt ny scraper fra et nytt bibliotek. - ligger fortsatt greit an

- Spør Raymond om hvordan dere kan måle suksess ut ifra hva dere har hentet på dark web? Skal det gå på lenker, mengder data, eller annet? Hva skal på resultat delen?

- Eksempler fra filene, hva har søkeroboten hentet ut?

21.04.23 – Statusmøte med veileder

Deltakere

- Erjon Zoto (veileder)
- Alle gruppemedlemmer

Agenda

- Metode
- Struktur
- Litteraturstudie
    - Tradisjonell eller systematisk?
        - Hvordan skal vi skrive at vi har gjort det? tabeller osv.?
        - Skal vi ha med søketreff osv. i tabellformat?
- Kildehenvisning
- Statistikk

Gjennomgang av kommentarer på 1. utkast:

- Mer utfyllende nå enn fra utkastet. Problemområde forskning i kildeliste. Beskrive mer rundt forskningen til Raymond, og hvorfor vår oppgave er viktig.
- Siden rapporten er på norsk, vil det også være vesentlig at målgruppen skal kunne norsk.
- Nytt utkast neste uke.
- Bestemme en spesifikk del av rapporten som skal jobbes på til neste uke, så skal det gås gjennom neste uke.
- Passe på å ikke repetere for mye.
- Lete etter andre bilder som representerer krypteringslag nåværende figur 2.2
- Så lenge all tekst på bildet er lesbart kan man gjøre bilder mindre.

Metode:

- Forklaring av flytdiagrammene.

Struktur:

- Usikre på plassering av DDOS-delen. Passer kanskje bedre innenfor bakgrunn/historie, hvert fall ikke under teknisk oppbygning. Mer egnet i diskusjonsdelen, siden dette ikke har vært planlagt i prosjektplanen.
    - En liten del i teori-delen, så referere videre til diskusjons-delen.
    - En begrensning/utfordring, så anbefales å nevne der.
- Ny struktur. Ha Tor over søkerobot delen. Litt mer komplisert med den nye strukturen.
    - Beholde den gamle strukturen.
- Søkerobot-kilde?
    - Kort introduksjon på hva en søkerobot gjør.
- Ikke nødvendig å ha bilder over spørreskjemaet. Mulig ha det som vedlegg, kanskje dette fjernes etter hvert.
    - Tabeller over spørsmål, istedenfor bilder og vedlegg.

Litteraturstudie:

- Ikke en ordentlig litteraturstudie som er blitt gjort. Kan egentlig ikke kategoriseres som tradisjonell eller systematisk.
    - Skriv litt rundt prosessen for hvorfor vi ikke har valgt å ikke utføre en «litteraturstudie».
- Skriv hva vi har gjort, og hva vi har søkt etter.
- Beskrive at oppgaven ikke er en litteraturstudie eller en utviklingsoppgave.
    - Skrive at det er en blanding, og nevne at litteraturstudien er blitt gjort gjennom en valgt søkemotor, og hvilke søkeord som er blitt brukt.
    - Beskriv en arbeidsfordeling rundt litteraturstudie og utvikling av søkeroboten.
- Videre analyse, beskriv hva andre forfattere mener, og diskutere hva VI mener om dette.
- Beskrive resultatene av en søkestrategi.
    - Med parametere, som år, søkeord og kilder.

Kildehenvisning:

- Siterer det første momentet vi bruker kilden, spesielt ved bruk av tall og målinger.
- På en måte en temasetning, som refereres til videre. Slik at man har ett sitat og videre henviser til det uten å henvise til kilder. Tror det var det som Raymond mente.

Statistikk:

- Hvis det ikke er statistikk, må vi bare skrive det.
- Blir vanskelig for Erjon å få tilgang på Statista, han må da kontakte økonomi og få tilgang på den måten. Erjon skal høre med Erik.

Forskningsspørsmål:

- Hva Tor handler om? (Litteraturstudie)
- Hvordan er det mørke nettet oppfattet? (Spørreundersøkelse)
- Hvordan kan man hente ut informasjon/data fra det mørke nettet? (Søkerobot)

Spørreundersøkelse:

- Burde sjekke validitet for denne.
- Vanskelig å bruke denne effektivt, og hvis vi ikke får brukt denne bra, kan den sløyfes.
- Hvis litteraturen støtter folket eller motsatt?
- Mulig at spørreundersøkelsen kunne vært sin egen oppgave.
- Underbygge legaliteten til Tor og det mørke nettet.
- Begrunne prosessen bak utsendingen av undersøkelsen.
- Analysere dataene fra undersøkelsen.
    - Begrunne og skrive rundt svarene.
- Metode, diskusjon og (videre arbeid)

Tittel:

- Det mørke nettet???
- Kommer tilbake til dette

02.05.23 – statusmøte med veileder

Deltakere

- Erjon Zoto (veileder)
- Thugitha Kanavathi
- Mats D. Jensen

Agenda

Spørsmål fra Marte

- Introduksjon – Bakgrunn, siste avsnitt om kvalitativ og kvantitativ metode. Er det vi skriver der korrekt for vår oppgave?
    -

- Første forskningsspørsmål, intensjon/hensikt/formål? Evt noe annet?

- Kan vi droppe å ha hypoteser siden vi har endret forskningsspørsmålene?

- Se på resultatmål/effektmål - hva skriver vi om vi ikke har oppfylt disse? Skal vi endre dem til noe vi har fått til? (For eksempel Resultatmål 1 om database, det har vi vel ikke?) og crawler med nøkkelord?
    - Så lenge vi har forsøkt å nå målet så kan vi skrive dette i diskusjonsdelen, at vi har prøvd, men at vi endret målet etter avtale med oppdragsgiver.

- Hentet fra kommentar i LaTeX: Martin: Høres lurt ut, kanskje høre med Erjon om vi skal ta med litteraturfunn som vi leser og som går gjennom kriteriene, men som vi ikke tar med pga relevansen for oppgaven?
    - Skrive en liten liste på det og HVORFOR vi ikke har brukt artiklene

- Skal vi diskutere læringsmålene i diskusjonskapitlet? Om det ble oppnådd, hva vi gjorde for å oppnå? Eller blir det en del av konklusjonen?
    - Bør diskuteres i konklusjonsdelen

- Vi har vel egentlig underveis kommet frem til at vi skal belyse at Tor også brukes av de med gode hensikter og at det i mange situasjoner er veldig viktig. Det har vi vel ikke med i avgrensing, problemområde? Kommer på teori, men kommer det litt brått? Burde vi ha noe mer om det i introduksjonen?
    - Thugitha har begynt å skrive litt mer under problemområde. Ref siste avsnitt under problemområde. NB!! Må skrives om!!

- Teknisk oppbygging: Inngangsnoder og formel for sannsynligheten om angriper. Tilfører det noe for leseren, eller forvirrer det?
    - Ikke nødvendig å ha med

- Spørre hvor mye som skal forklares i glossary? Vanskelig å vite hvor vi skal "legge lista".
    - Kommer an på målgruppen, hvis målgruppen er IT-folk så bør ikke ord som ruter osv. defineres, men dersom målgruppen er alle potensielle brukere av Tor så bør alt defineres.

- DDoS i teori: Kutte første avsnitt (og heller forklare ddos i glossary) og flytte avsnitt 2 til diskusjon?

- Metode, håper virkelig han har litt kommentarer der om vi er på rett vei osv.
    - Vi er på rett spor, men kan gjøre det mer om til et sammendrag dersom vi ikke klarer å skrive alt i detalj.

- På resultat: Skrive kortere om hvert funn? Eller er det ok?

Omskrive siste avsnitt 1.2

1.3 Kun ha oppgaven fra Raymond. Er Datamengder fortsatt relevant? Spør Raymond.

1.4.1 Flette sammen spm. 1 og 2. ---- Forskningsspørsmål trenger ikke være spesifikke, hold det generelt.
Hva er det mørke nettet- hva er formålet - heller ha underspørsmål fra dette.
Få frem hva det mørke nettet er, men at vi holder oss til Tor. Beskriv forskningsspørsmålene få mer tekst inn her.
Kan ha en paragraf for begge, eller dele dem opp.

Trenger ikke ha med Hypoteser. Kan ha hvis vi vil. Hvis vi har det, så må vi forklare godt rundt det. Spm. 1 fungerer ikke så bra, heller ha: "Det mørke nettet har både positive og negative funksjoner". Definer det på en bedre måte om det skal brukes.

1.6.1 knytt sammen forskningsspørsmål og effektmålene.

Unngå å bruke ordet "rykte" hold oss til kun fakta.

Prøv å få mer beskrivende titler. - i subsections også. I stedet for Russland - "Blokkering av internett"

Tjenestenektangrep under 2.2?

2.5 - flere eksempler på søkeroboter?

Litteraturstudie – trenger ikke være så spesifikk på alt, kan oppsummere der vi kan, så slipper vi å bruke så mye tid på det. Fyll ut, så ser vi senere på hvordan vi gjør det videre.

Søkeroboten kjører fra forskjellige maskiner, og ulike startpunkt, forskjellige resultater ut fra dette? Vis skjermbilder av hva som er hentet ned (Aner ikke hvordan – finn noe som ser bra ut)

Erjon: Kan det være en idé å ha med skjermbilder av sider fra Tor for å underbygge påstanden om at Tor blir brukt til gode formål. For eksempel bilde av whisterblower-sider.

Litteraturstudiet bør ikke ha noen resultater, dette blir skrevet om i teori-delen. Det er mulig å forklare dette i metode-delen.

VIKTIG!!!!! Vise til data som har blitt høstet av søkeroboten.

Trenger ikke ha med alternativer i tabellene til spørreundersøkelsen, Få alle i samme tabell. - seksjoner som kategoriserer spørsmålene.

Tallene fra spørreundersøkelse kan brukes for å forklare kvantitativ metode – i kap. 3

Grunnet dårlig resultat valgte vi å ikke implementere nøkkelord for søkeroboten.

11.05.23 – Krisemøte internt i gruppen

Deltakere

- Mats Jensen
- Martin Hyldmo
- Marte Jørgensen
- Thugitha Kanavathi

Agenda

- Hva er tittelen på oppgaven?
- Hva er forskningsspørsmålene?
- Ting som må gjøres

Hva er tittelen på oppgaven?

- Placeholder: Et dypdykk i TOR
- Anonym kommunikasjon på det mørke nettet - et dypdykk i TOR

Hva er forskningsspørsmålene?

- Heller ha problemstilling og problemformulering?

- Forslag: flytte avgrensninger opp slik at det blir under oppgavedefinisjonen, og skrive i avgrensinger at "som nevnt i oppgavedefinisjonen vil denne rapporten fokusere på Tor"

To do list

- Skrive om problemområde
- Legge til mer i bakgrunn sånn at det blir en sammenheng mellom dette og resten av oppgaven
- Skrive problemformulering
- Omformulere problemstillingen
- Metode om litteraturstudie må fullføres
- Resultatet av datahøstingen må fullføres
- Hele diskusjon
- Hele konklusjon

Til presentasjonen:

- Finne ting vi kunne ha gjort annerledes og begrunne hvorfor ting ble som det ble

16.05.23 – Statusmøte med veileder

Deltakere
- Erjon Zoto
- Marte Jørgensen
- Mats Jensen
- Martin Hyldmo
- Thugitha Kanavathi

Agenda
- Gå gjennom siste utkast

Referat:

Kommentarer:
1.71. Effektmål - før eller etter høsting: Kronologisk rekkefølge.

Problemstilling/problemformulering: Det er bra, trenger ikke bytte tilbake til å ha forskningsspørsmål.

Metode: langt kapittel. Dele opp i hvert sitt kapittel, søkerobot/spørreundersøkelsen. Fra metode til resultat måtte vært med i hvert kapittel for disse to evt. Får gjøres kun hvis det blir tid. La teori stå?

Metode innhold: OK.

Resultat: Kanskje 4.15 og 4.15 (figur) viser nesten det samme. Det som vises fra crawler sin del og mapper. Kunne kanskje skille. Kanskje 4.16 kan vise mer detaljer. ER GJORT. Erjon fornøyd nå.
Ta en vurdering på om 4.16 skal bort!

Vær forsiktig med bilder som er snapshots, det kan hende det ikke er så bra til å visualisere – 4.12. Kanskje bare skrive og ikke bilder. Figur 4.13: Den som viser totale størrelse på harddisken, trenger ikke! Bare skrive det.

Resultat/diskusjon: Søkerobot: Kjørtes først på det åpne nettet – fant veldig mye data på kort tid. På det mørke nettet mye vanskeligere. Vi har en god del utfordringer vi har møtt på.
Annen måte, annen type crawler, annen bibliotek: Kan vi vurdere dette? Hvis det har fungert bra, skriv ned. Crawler kan både hente og laste ned -> Raymonds ønske.

Tabell 4.1: Største filen lastet ned, avbrutt og fil kan ikke åpnes. Så vi vet ikke helt hva det er.

Resultat: Bør vi vise mer av dataene vi har hentet, eller holder det vi har nå? Erjon skulle gjerne hatt litt mer informasjon om innholdet. Hva slags typer data er det, er det stjålet data? Organisert kriminalitet? Video, bilder? Kopi av noen sin PC. Være mer detaljert. For eksempel 70 % er knyttet til hacking osv. Eller knytte opp mot resultatene i spørreundersøkelsen. Vi har leita etter ulovlig ting på nettet, så blir ensidig kanskje?

Diskusjon: Erjon ser på den etter lunsj, vi sender siste versjon vi har!

For stor oppgave, ikke helt klart å avgrense oss ordentlig i starten. Tor er spennende, og man vil vite mer og mer. Når man sier man skal høste data fra Tor og mørke nett, det som forventes er at man vet hva det handler om, hva de snakker om først. Denne dataen vi har lastet ned, skrive mer detaljer er lurt.

Det er alltid viktig å knytte sammen de forskjellige delene. For eksempel vi gjorde spørreundersøkelsen for å forstå hva befolkningen mener om det mørke nettet. Søkerobot gir oss info om dataene som er der, og dette må knyttes sammen på et vis. Man må gå frem og tilbake og referere til deler vi har med slik at vi får flettet alt sammen.

Spørreundersøkelse: Viktig å begrunne hvorfor vi har valgt å ha den med, det var ikke med i oppgavebeskrivelsen til R. Vi gikk forbi og prøvde å tilføre noe mer, gjøre noe mer. Hvorfor, hva gir den oss.

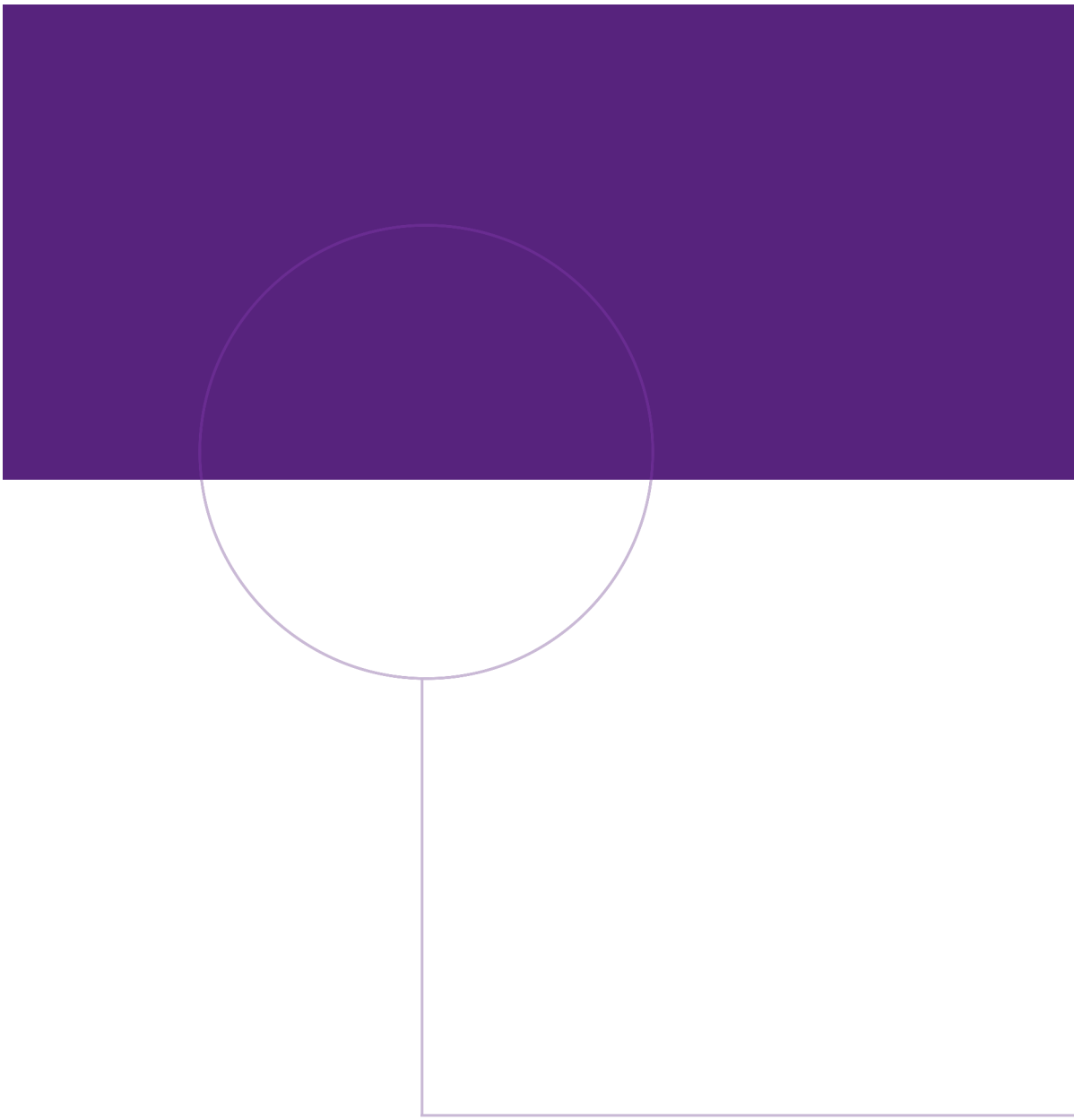
Spørreundersøkelsen har kanskje ikke så mye å gjøre med søkerobot. Men prøve å knytte sammen, for eksempel at befolkning har negativ assosiasjon og data lastet ned er bare kriminell aktivitet.

Skriv mer sånn at spørreundersøkelsen blir integrert i hele studiet, så man får følelsen av at det er en naturlig del av oppgaven.

Spørreundersøkelse ikke en del av Raymond sin oppgavebeskrivelse. Dette er noe VI bestemte. Vi prøvde å gjøre noe mer enn det det var spurt etter. Knytte sammen fra metodedel, men mest viktig å diskusjonen. "Vi tenkte det var nyttig fordi bla bla, og nå ser vi at det var det, evt ikke" Forklar hvorfor og hvordan det henger sammen. Viste den det vi trodde?

Hovedmålet i skriving: Skriv formelt, tydelig og ikke for mye eller for lite.

Figur spørreundersøkelse: Spørsmålsinnhold: Hva er din tekniske bakgrunn. Klippe den fra bilde og bruke som tittel fra figur. Jo mindre tekst jo bedre under bilder.



NTNU

Kunnskap for en bedre verden