Viivi Järvinen

Tiedonkeruu penetraatiotestauksessa

Tietotekniikan kandidaatintutkielma

17. toukokuuta 2022

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Viivi Järvinen

Yhteystiedot: viivi.jarvinen96@gmail.com

Ohjaaja: Tytti Saksa

Työn nimi: Tiedonkeruu penetraatiotestauksessa

Title in English: Information gathering in penetration testing

Työ: Kandidaatintutkielma

Opintosuunta: Kaikki opintosuunnat

Sivumäärä: 25+0

Tiivistelmä: Tämän tutkielman aihe on tiedonkeruu penetraatiotestauksessa. Penetraatiotestaukseen tutustuminen on helppo aloittaa ensimmäisestä vaiheesta eli tiedonkeruusta. Tutkielmassa tutkitaan, mitä työkaluja tiedonkeruuseen käytetään. Lisäksi selvitetään työkalujen eroja. Tutkimuksessa selvisi, että menetelmien näkyvyyksissä ja käytettävyyksissä on eroja. Tutkimus antaa lukijalle hyvät pohjatiedot tiedonkeruuseen.

Avainsanat: Tiedonkeruu, penetraatiotestaus, tiedustelu, google hakkerointi, social engineering

Abstract: The topic of this thesis is information gathering in penetration testing. Familiarization with penetration testing is easy to start from the first step which is information gathering. The focus of this study is to map out the tools that are used in information gathering. The goal is to find out their purpose and how they differ from each other. This study showed that there are differences in visibility and usability. This thesis is a good way to get to know the information gathering phase.

Keywords: Information gathering, penetration testing, reconnaissance, Kali Linux, social engineering, google hacking

Kuviot

Taulukot

Sisällys

1 Johdanto

Tässä kandidaatintutkielmassa tutkitaan penetraatiotestauksen ensimmäistä vaihetta, joka tunnetaan tiedonkeruuna. Penetraatiotestaukseen tutustuminen on hyvä aloittaa ensimmäisestä vaiheesta. Tiedonkeruu on nimensä mukaisesti tiedon keräämistä kohteesta. Kohde voi olla esimerkiksi yritys. Tiedonkeruun suorittamiseen on monia tapoja. Tässä tutkielmassa selvitetään tiedonkeruun yleisimmät työkalut sekä tavat. Lisäksi tutkitaan eri tiedonkeruutapojen eroja. Tietoa voidaan kerätä internetistä, mutta tietoa voidaan pyrkiä saamaan myös muista lähteistä. Tietoa voidaan kerätä vaikkapa kohteen järjestelmistä tai henkilöstöstä. Tiedonkeruu penetraatiotestauksessa on tärkeää, koska se antaa informaatiota kohteesta. Tietoturvakartoituksella tietoturva-asiantuntija tietää, mitä työkaluja testauksessa kannattaa käyttää (Broad, Bindner ja Bindner 2013, s.89-90). Tiedonkeruuvaihe pohjustaa penetraatiotestauksen tulevia vaiheita (Engebretson ja Kennedy 2013, s.20). Tämä tarkoittaa sitä, että kerättyä tietoa tullaan käyttämään testauksen myöhäisemmissä vaiheissa kuten skannauksessa. Broad, Bindner ja Bindner (2013) tarkentavat kirjassaan, että tiedonkeruussa kohdetta ei yritetä vielä hakkeroida. Engebretson ja Kennedy (2013) jakavat kirjassaan testauksen vaiheet neljään osaan ja tutkielma mukailee tätä jakoa.

Tutkielman aiheesta löytyy paljon akateemista kirjallisuutta ja tekijöiden tavat lähestyä tiedonkeruuta eroavat toisistaan. Kirjoittajat suosivat joitakin työkaluja enemmän kuin toisia. Jokaisella on omanlainen tapa suorittaa tiedonkeruu. Tutkielmassa käydään läpi näitä erilaisia tiedonkeruun tapoja. Lisäksi tutkielmassa pyritään selvittämään, onko esimerkiksi HTTrack parempi tapa skannata verkkosivu kuin Wget. Tutkielmassa pohditaan, miten erilaiset tiedonkeruutavat eroavat toisistaan käytettävyydeltä sekä näkyvyydeltä.

Tutkimuksen tarkoituksena on antaa kattava kuva tiedonkeruusta ja sen työkaluista. Tutkielman toisessa luvussa avataan penetraatiotestauksen sekä tiedonkeruun käsitteitä. Tiedonkeruun tapoja on monenlaisia ja siksi on tärkeää vertailla niitä keskenään. Kolmannessa luvussa kartoitetaan tiedonkeruun erilaisia tapoja ja neljännessä luvussa näitä tapoja vertaillaan. Lopputuloksena on tutkielma tiedonkeruun tavoista ja näiden tapojen eroista käytettävyyden sekä näkyvyyden näkökulmasta. Tutkielma kokoaa ja vertailee materiaalia sellaiseen muotoon, että aiheeseen on helppo tutustua.

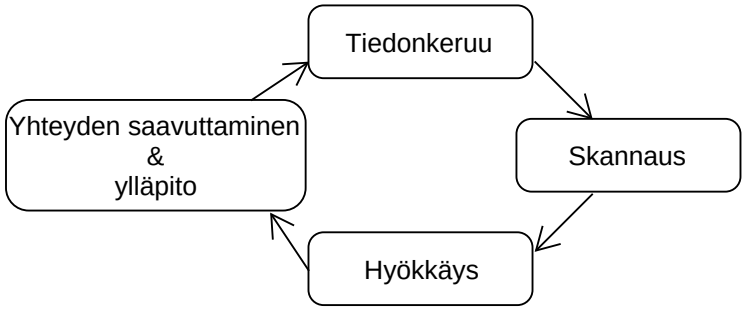
2 Tiedonkeruu osana penetraatiotestausta

Tässä osiossa esitellään tiedonkeruun ja penetraatiotestauksen keskeisiä käsitteitä. Penetraatiotestaus on yksi tietoturva-arvioinnin tapa (Allen, Tedi ja Shakeel 2014). Penetraatiotestaus sisältää erilaisia vaiheita, jotka voidaan määrittää eri tavoin penetraatiotestaajasta riippuen. Tässä tutkielmassa tiedonkeruun vaiheet on jaettu tiedonkeruuseen, skannaukseen, hyökkäykseen ja yhteyden ylläpitoon.

2.1 Penetraatiotestaus

Tiedonkeruuseen liittyvä tärkeä käsite on penetraatiotestaus, koska tiedonkeruu on osa penetraatiotestausta. Penetraatiotestaus on rikollisen hakkerin käytöksen mallintamista tietoturvan testaamiseksi (Broad, Bindner ja Bindner 2013, s.3). Kuten aiemmin mainittiin, tiedonkeruun rooli testauksessa on alustaa muita vaiheita. Alustuksella tarkoitetaan sitä, että kerättyä tietoa hyödynnetään testauksen muissa vaiheissa sekä luodaan testaajalle kattava kuva kohteesta. Engebretson ja Kennedy (2013) määrittelevät penetraatiotestauksen luvalliseksi hakkeroinniksi. Testauksen tarkoitus on kartoittaa kohteen tietoturva etsimällä mahdollisia aukkoja kohteen suojauksessa (Allen, Tedi ja Shakeel 2014, s.51-52). Tarkoituksena on parantaa tietoturvaa löytämällä heikkoudet ennen kuin luvaton taho pääsee niitä hyödyntämään.

Tässä tutkielmassa penetraatiotestaus jaetaan neljään vaiheeseen kuvion 1 mukaan. Tiedonkeruu on näistä vaiheista ensimmäinen. Loput vaiheet ovat skannaus, hyökkäys sekä yhteyden saavuttaminen ja ylläpito (Engebretson ja Kennedy 2013, s.1). Skannaus koostuu porttien sekä kohteen heikkouksien skannaamisesta, kun taas hyökkäyksen rooli penetraatiotestauksessa on päästä hyödyntämään löydettyjä sisäänkäyntejä ja saada kohteen järjestelmät hallintaansa (Engebretson ja Kennedy 2013, s.1). Hyökkäyksen ylläpito Wilhelmin mukaan tarkoittaa sisäänpääsyn säilyttämistä kohteeseen (Wilhelm 2013). Luvattomat hakkerit ovat yleensä kiinnostuneita kohteen pitkäaikaisesta hakkeroinnista ja yhteyden ylläpidosta, joten se on hyvä sisällyttää penetraatiotestaukseen (Engebretson ja Kennedy 2013, s.1). Penetraatiotestaus voidaan jakaa myös useampaan osaan. Testaukseen voitaisi sisällyttää esimerkiksi tulosten raportointi, kuten Wilhelm (2013) on kirjassaan tehnyt.



Kuvio 1. Penetraatiotestauksen vaiheet. Mukailtu lähteestä (Engebretson ja Kennedy 2013, s.19)

2.2 Tiedonkeruu

Tiedonkeruu on nimensä mukaisesti tiedon keräämistä kohteesta. Tiedonkeruuta voidaan toteuttaa esimerkiksi internetissä olevilla hakukoneilla (Broad, Bindner ja Bindner 2013, s.87). Tiedonkeruuta voidaan myös kutsua tiedusteluksi (engl. reconnaissance). Tiedonkeruun tarkoitus penetraatiotestauksessa on antaa mahdollisimman kattava kuva kohteesta sekä sen järjestelmistä (Wilhelm 2013, s.151). Jo kerättyä tietoa voidaan myös käyttää toisen tiedon keräämiseen ja tällöin voidaan saadaan enemmän informaatiota kohteesta.

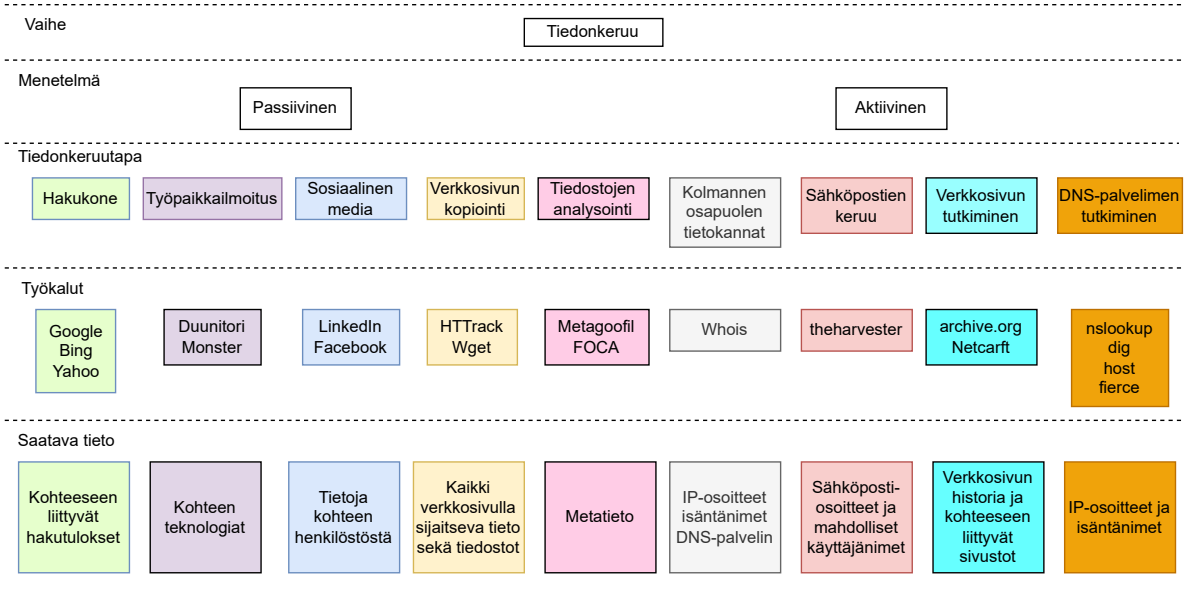
Tiedonkeruu voidaan toteuttaa aktiivisella tai passiivisella tiedustelulla (Engebretson ja Kennedy 2013, s.22). Aktiivinen tiedonkeruu on yhteyden muodostamista kohteeseen ja passiivisessa yhteyttä ei muodosteta (Engebretson ja Kennedy 2013, s.22). Toisin sanoen passiivisessa tiedustelussa kohde ei ole tietoinen tiedonkeruusta, kun taas aktiivisessa tiedonkeruussa kohde voi havaita penetraatiotestaajan toimet. Aktiiviseen tiedonkeruuseen liittyy vahvasti DNS, joka on nimipalvelujärjestelmä. Nimipalvelujärjestelmän tarkoitus on muuttaa verkkotunnukset IP-osoitteiksi. DNS-palvelinta käsitellään tarkemmin aktiivisen tiedonkeruun osiossa.

Tiedonkeruuseen on monia erilaisia työkaluja, joita tutkielmassa tullaan käsittelemään. Tällaisia työkaluja ovat esimerkiksi hakukoneet ja yksi tunnetuimmista hakukoneista on ominaisuuksiltaan monipuolinen Google. Google on ominaisuuksiltaan niin laaja, että tiedonkeruu googlesta on saanut oman nimensä: Google-hakkerointi. Google-hakkerointi tarkoittaa

Google hakukoneen käyttöä tietojen keräämiseen. Tiedonkeruu tapahtuu Googlen valtavasta tietokannasta käyttäen hakuun sopivia kyselyitä (Broad, Bindner ja Bindner 2013, s.97). Hakukoneiden lisäksi löytyy paljon erilaisia työkaluja, joita tullaan käsittelemään tiedonkeruun menetelmissä.

Kali Linux on penetraatiotestaukseen luotu Linux-käyttöjärjestelmä, joka sisältää testaukseen tarvittavia työkaluja ("Kali" 2022). Kali Linux on tehty penetraatiotestausta varten (Engebretson ja Kennedy 2013, s.7). Kali Linuxista löytyy laaja kattaus työkaluja, joilla voidaan suorittaa tiedonkeruu. Ohessa vielä kuvio 2 tiedonkeruutavoista, työkaluista ja tiedosta, jota yritetään kerätä. Näihin syvennytään seuraavassa luvussa. Kuvio 2 ei ole täysin kattava ja esimerkiksi seuraavassa kappaleessa käsiteltävä käyttäjän manipulointi puuttuu.

Erilaisten työkalujen lisäksi tiedonkeruu voidaan toteuttaa käyttämällä hyväksi inhimillistä tekijää. Kyseistä menetelmää kutsutaan käyttäjän manipuloinniksi (engl. social engineering). Menetelmässä on kyse inhimillisen tekijän hyödyntämisestä tietojen saamiseksi (Engebretson ja Kennedy 2013, s.48). Penetraatiotestauksessa käyttäjän manipulointi tarkoittaa ihmisten välistä kanssakäymistä, jonka tarkoituksena on saada tietoja toiselta osapuolelta (Patel 2013, s.6). Patel (2013) kirjassaan jakaa kanssakäymisen ihmis- ja tietokonepohjaiseen käyttäjän manipulointiin.



Kuvio 2. Tiedonkeruun menetelmät. Saatava tieto on tietoa, joka saavutetaan onnistuttaessa.

3 Tiedonkeruun menetelmät

Tiedonkeruun tapoja on monenlaisia. Kaikki tiedonkeruun yritykset eivät aina onnistu ja siksi on käytettävä useaa eri tapaa tiedon keräämiseen. Mikäli erilaiset tiedonkeruun menetelmät halutaan erotella, ne voidaan jakaa passiiviseen ja aktiiviseen tiedonkeruuseen. Passiivinen tiedonkeruu on paljon nimettömämpää kuin aktiivinen. Aktiivinen tiedonkeruu on puolestaan näkyvämpää kuin passiivinen. Kummallakin tiedonkeruun tavoilla saadaan tärkeää tietoa kohteesta. Näiden lisäksi joskus voidaan puhua fyysisestä tiedustelusta johon kuuluu esimerkiksi käyttäjän manipulointi (Shimonski, Bishop ja Zenir 2014, s.64).

Tiedonkeruun tarkoituksena on oppia tuntemaan kohde (Engebretson ja Kennedy 2013, s.53). Engebretson ja Kennedy (2013) kertovat kirjassaan, että tiedonkeruussa pyritään muodostamaan kokoelma kohteen tietoja. Kohteen IP-osoitteet tulee olla listattuna tähän kokoelmaan, koska niitä tullaan käyttämään skannauksessa (Engebretson ja Kennedy 2013). IP-osoitteet skannataan avoimien porttien toivossa (Baloch 2015). Nämä portit tarjoavat mahdollisen sisäänpääsyn kohteeseen, jonka vuoksi IP-osoitteiden kartoittaminen on tärkeää. Engebretson ja Kennedy (2013) kertovat kirjassaan, että kohteen DNS-palvelin pitää sisällään kohteelle kuuluvat IP-osoitteet ja isäntänimet. Tiedonkeruussa pyritään selvittämään kohteen DNS-palvelimet, jotta saadaan haltuun niiden sisältämät IP-osoitteet. DNS-palvelimelle voidaan yrittää tehdä DNS-vyöhykkeen siirto, jota tullaan käsittelemään luvussa 3.2. IP-osoitteiden lisäksi halutaan myös muita tietoja, jotta osataan suunnitella tulevien vaiheiden toteutus. Näiden tietojen saavuttamiseen käytetään tiedonkeruuta. Erilaisten tietojen keräämiseen käytetään erinäisiä työkaluja. Kerättyjä tietoja voidaan hyödyntää tiedonkeruussa tai käyttää penetraatiotestauksen myöhäisemmissä vaiheissa.

3.1 Passiivinen

Passiivisiin tiedonkeruumenetelmiin kuuluu kohteen web-läsnäolon kartoitus (Wilhelm 2013, s.153). Lähteen mukaan läsnäolon kartoitus tapahtuu hakemalla tietoa esimerkiksi sosiaalisesta mediasta, työntekijöiden henkilökohtaisilta sivustoilta, työpaikkailmoituksista sekä käyttämällä hakukoneita. Internetistä saatava tieto, joka ei jätä lokitietoja kohteelle, on pas-

siivista tiedonkeruuta. Wilhelmin mukaan passiivista tiedonkeruuta voi myös suorittaa kyselemällä kolmannen osapuolen nimipalvelimelta tietoja kohteen verkkotunnuksen avulla (Wilhelm 2013).

Passiivinen tiedonkeruu voidaan aloittaa hakukoneilla, joilla etsitään tietoa kohteesta. Aluksi voidaan etsiä kohteen verkkosivut ja verkkotunnus. Kun kohteen verkkosivu on löydetty, voidaan verkkosivusta tehdä paikallinen kopio. Kopioimiseen tarkoitettuja työkaluja ovat HTTrack, Linuxissa olevaa Wget ja Website Ripper Copier (Baloch 2015, s.54-55). Kopioitu verkkosivu sisältää kaikki sivustolla olevat sivut sekä niiden sisällön. Näin ollen verkkosivun kopiointi tarjoaa paljon informaatiota kohteesta. Vaikka verkkosivun kopiointi voidaan havaita, Engebretson ja Kennedy (2013) täydentävät kirjassaan, että kohteen sivuston kopioimista pidetään passiivisena tiedonkeruuna. Heidän mukaansa kopiointi on passiivista, koska tarkoituksena on päästä tutkimaan sivustoa jättämättä lokitietoja kohteelle. Toisin sanoen tietoa kerätään sivustolta offline-tilassa. Tämä vähentää penetraatiotestaajan näkyvyyttä kohteen lokitiedoissa.

Verkkosivujen kopioimisen lisäksi tietoa voidaan saada tutkimalla verkkosivuilla tapahtuneita muutoksia. Archive.org on sivusto, joka sisältää tiedot vuosien aikana sivustolla tapahtuneista muutoksista. Archive.org näyttää, mitä sivustolta on poistettu, päivitetty ja lisätty. Poistetut tiedostot voivat sisältää arkaluontoisia tietoja, joita sivustolle on saatettu vahingossa vuotaa. (Wilhelm 2013, s.157-158)

Kohteen verkkosivun tutkimisen lisäksi monet hakukoneet tarjoavat paljon tietoa passiivista tiedonkeruuta suorittavalle. Tällaista hakukoneilla tehtävää tiedonkeruuta voidaan toteuttaa työpaikkahakukoneilla kuten Duunitori ja Monster. Työpaikkailmoitukset voivat kertoa minkälaista teknologiaa kohde käyttää (Broad, Bindner ja Bindner 2013, s.99). Kirjassaan Engebretson ja Kennedy (2013) kertovat, että penetraatiotestaaja voi tehdä päätelmiä työpaikkailmoituksen sisältämistä teknologioista. Muita hakukoneita ovat Googlen lisäksi Bing ja Yahoo. Sosiaaliset mediat kuten LinkedIn ja Facebook puolestaan tarjoavat paljon tietoja ihmisistä ja kohteen avainhenkilöistä.

Edellisissä kappaleissa käsiteltiin internetistä saatavaa tietoa. Internetissä on hyviä työkaluja passiivisen tiedonkeruun suorittamiseen kuten verkkoselaimessa toimiva Netcraft. Ba-

loch (2015) kertoo kirjassaan, että Netcraft on tietokanta, joka sisältää tietoja verkkosivuista. Netcraftillä voidaan selvittää kohteen verkkopalvelimen käyttöjärjestelmä (Baloch 2015, s.63). Engebretson ja Kennedy (2013) mainitsevat kirjassaan Netcraftin palauttavan kaikki sivut, jotka sisältävät hakuun käytetyn hakusanan. Heidän mukaansa sivustoraportti paljastaa sivuston IP-osoitteen ja DNS-palvelimen. Löydökset voidaan lisätä IP-osoitteiden kokoelmaan.

Kali Linuxissa valmiina oleva theharvester on työkalu, jolla hakukoneista voidaan hakea sähköposteja sekä aliverkkotunnuksia (Engebretson ja Kennedy 2013, s.33). Työkaluna theharvester eroaa jo käsitellyistä työkaluista siten, että sitä ei voi käyttää verkkoselaimella. Theharvester pitää asentaa tai olla asennettuna laitteeseen, jotta sähköpostihaku voidaan suorittaa. Sähköpostien keruu tapahtuu käyttämällä hyväksi julkista tietoa, joten kohteeseen ei luoda yhteyttä (Baloch 2015, s.69). Theharvester voi käyttää sähköpostiosoitteiden hakuun hakukoneita kuten Google tai LinkedIn (Engebretson ja Kennedy 2013, s.32). Sähköpostiosoite voi sisältää työntekijöiden mahdollisia käyttäjänimiä, joilla voidaan hakea lisää tietoa internetistä (Engebretson ja Kennedy 2013, s.32).

Seuraava käsiteltävä työkalu on Metagoofil. Theharvesterin tapaan Metagoofil on ohjelma, joka tulee asentaa omalle laitteelle. Metagoofil kerää metatietoa (Engebretson ja Kennedy 2013, s.44). Metatieto on tietoa tiedosta. Tässä yhteydessä metatieto tarkoittaa informaatiota esimerkiksi tiedostosta. Tiedoston metatieto voi sisältää tietoa vaikkapa tiedoston tyypistä, omistajasta, tiedostosijainnista sekä viimeisestä muokkauspäivästä (Engebretson ja Kennedy 2013, s.44). Tiedostonsijainti voi tarjota hyökkääjälle tietoa hakemistosta. Edellä mainittujen tietojen lisäksi metatieto voi sisältää myös muita tietoja riippuen tiedostosta. Engebretson ja Kennedy (2013) tuovat kirjassaan ilmi, että Metagoofil etsii Internetistä kohteelle kuuluvia tiedostoja ja analysoi niitä. Foca on myös metatietoa analysoiva työkalu, joka etsii tiedostoja Googlesta, Yahoosta ja Bingstä lataamatta niitä (Baloch 2015, s.68).

Seuraavaksi käsiteltävä työkalu Whois tulee esiin useissa käsitellyissä lähteissä, joten sitä voidaan pitää merkittävänä passiivisen tiedonkeruun työvälineenä. Whois on tietokanta, joka sisältää laajasti tietoa netissä olevista sivustoista (Baloch 2015, s.55). Whois-haulla voidaan saada selville kohteen IP-osoitteet, isäntänimet (eng. hostnames), DNS-palvelimet ja avainhenkilöiden yhteystiedot (Engebretson ja Kennedy 2013, s.34). Engebretson ja Kenne-

dy (2013) kehottavat kiinnittämään erityistä huomiota DNS-palvelinten selvittämiseen. Palvelimet sisältävät paljon tietoa, joten niihin on syytä kiinnittää huomiota. Whois-työkalulla haetaan tiedot käyttäen verkkotunnusta. Whois voi haun tuloksena antaa viittauksen toiseen verkkotunnukseen ja uusi Whois-haku voidaan toteuttaa uudella verkkotunnuksella (Engebretson ja Kennedy 2013, s.36). Whois-hauilla pyritään selvittämään mahdollisimman monta verkkotunnusta, jotka voidaan muuttaa IP-osoitteiksi. Tietojen keruu tapahtuu Whois-tietokannasta eikä suoraan kohteelta (Allen, Tedi ja Shakeel 2014, s.85). Haku on passiivista, kun kyseessä on kolmannelta osapuolelta saatava tieto. Kun DNS-palvelimet ovat listattuna nimin, voidaan ne muuttaa IP-osoitteiksi käyttämällä hostia (Engebretson ja Kennedy 2013, s.35). Allen, Tedi ja Shakeel (2014) mukaan hostilla voidaan myös tehdä käänteinen haku syöttämällä IP-osoite ja tuloksena saadaan isäntänimi.

Whoisin lisäksi lähteissä käsitellään paljon työkaluja nslookup ja dig. Engebretson ja Kennedy (2013) määrittelevät nslookupin seuraavasti: nslookup on työkalu, jolla voidaan suorittaa kyselyitä DNS-palvelimelle. Nslookup voidaan käynnistää komentoriviltä komennolla nslookup. Dig on toiminnaltaan samanlainen kuin nslookup (Baloch 2015, s.76). NsLookupin sekä digin kohdalla kannattaa olla varovainen, ettei kyselyitä tee suoraan kohteen nimipalvelimelle, koska silloin siirrytään passiivisesta tiedonkeruusta aktiiviseen. Sen sijaan tietoa kysellään julkisilta DNS-palvelimilta. Passiivisessa tiedonkeruussa keskitytään vain ei-valtuutetun tiedon kyselyyn. Mikäli nslookupille syöttää palvelin komentona kohteen palvelimen, on tiedonkeruu aktiivista. Ilman palvelimen määrittämistä nslookup käyttää oletusnimipalvelinta. (Wilhelm 2013, s.170-174)

Tässä luvussa käsiteltyjen työkalujen lisäksi on myös työkaluja, jotka suorittavat monta erilaista tiedonhakua. Penetraatiotestaajille on tarjolla tiedonkeruuseen niin sanottuja monitoimityökaluja. Allen, Tedi ja Shakeel (2014) esittelevät tällaisiksi työkaluiksi DMitryn sekä Maltegon. DMitry hakee IP-osoitteet, isäntänimet, verkkotunnuksen, aliverkkotunnukset, kohteen sähköpostiosoitteet ja lisäksi se suorittaa porttiskannauksen (Allen, Tedi ja Shakeel 2014, s.100-103). DMitry tosin suorittaa jo aktiivista tiedonkeruuta sekä skannausta. Allen, Tedi ja Shakeel (2014) kertovat, että Maltego kerää verkkotunnukset, IP-osoitteet, suorittaa Whois-haun ja kerää tietoa kohteen ihmisiin liittyen.

3.2 Aktiivinen

Kuten aiemmin on käsitelty, aktiivisessa vaiheessa luodaan yhteys kohteeseen. Aktiivisessa tiedonkeruussa kommunikoidaan kohteen verkon kanssa (Allen, Tedi ja Shakeel 2014, s.85). Wilhelm (2013) toteaa, että aktiivisessa tiedonkeruussa yleensä saavutetaan ajantasainen tieto. Aktiivisella tiedonkeruulla voidaan varmistaa passiivisessa tiedonkeruussa kerättyjen tietojen paikkansapitävyys (Wilhelm 2013, s.172). Baloch (2015) laskee aktiiviseen tiedusteluun myös porttien skannauksen, mutta tässä tutkielmassa skannausta ei lasketa tiedonkeruuvaiheeseen. Tämä tutkielma keskittyy aktiivisessa tiedonkeruussa lähinnä tietojen saamiseen DNS-palvelimelta. Myös käyttäjän manipulointia voidaan pitää aktiivisena tiedonkeruuna, mutta laajuutensa vuoksi se on käsitelty erillisenä osiona.

Kun passiivisella tiedonkeruulla on saatu mahdollisimman paljon tietoa, voidaan aloittaa aktiivinen tiedonkeruu saaduilla tiedoilla. Aktiivisen ja passiivisen tiedonkeruun tapoja saattaa olla vaikea erottaa, koska samoja työkaluja voidaan käyttää sekä passiiviseen, että aktiiviseen tiedonkeruuseen. Passiivisen tiedonkeruun osiossa esiteltiin työkalut dig, nslookup ja host. Nämä ovat passiivisen tiedonkeruun työkaluja, kunnes muodostetaan yhteys kohteeseen (Wilhelm 2013, s.167-173).

DNS-palvelin muuttaa verkkotunnuksia IP-osoitteiksi kuten luvussa 2.2 määriteltiin. Aktiivisessa tiedonkeruussa keskitytään kohteelle kuuluviin DNS-palvelimiin, koska ne sisältävät tietoja kohteen sisäisistä IP-osoitteista sekä isäntänimistä (Engebretson ja Kennedy 2013, s.39-40). Engebretson ja Kennedy (2013) määrittelevät kohteen IP-osoitteiden keräämisen yhdeksi tiedonkeruun päämääristä. IP-osoitteita käytetään skannauksessa eli penetraatiotestauksen toisessa vaiheessa. Jotta IP-osoitteet saadaan selville kohteen DNS-palvelimelta, pitää DNS-palvelinta kuulustella.

DNS-kuulustelu on yksi aktiivisen tiedonkeruun tapa. Passiivisen tiedonkeruun kappaleessa käsitellyt dig ja nslookup ovat DNS-kuulusteluun käytettyjä työkaluja. Engebretson ja Kennedy (2013) kertovat, että edellä mainituilla työkaluilla voidaan suorittaa DNS-vyöhykkeen siirto. Allen, Tedi ja Shakeel (2014) toteavat, että hostilla voidaan myös suorittaa DNS-vyöhykkeen siirto. Passiivisessa tiedonkeruussa kyseltiin tietoja kolmannen osapuolen DNS-palvelimelta. Erona passiiviseen tiedonkeruuseen onkin se, että aktiivisessa yhteys muodos-

tetaan suoraan kohteeseen.

DNS-kuulustelu voidaan toteuttaa DNS-vyöhykkeen siirrolla. Passiivisen tiedonkeruun vaiheessa on saatu kohteen IP-osoitteet selville ja nyt niitä käytetään aktiiviseen tiedonkeruuseen (Engebretson ja Kennedy 2013, s.40). Vyöhykkeen siirto tarkoittaa kaiken nimipalvelimella olevan tiedon siirtämistä (Broad, Bindner ja Bindner 2013, s.102). Vyöhykkeen siirto tapahtuu tekemällä DNS-tietokannasta kopio, joka siirretään halutulle palvelimelle (Allen, Tedi ja Shakeel 2014, s.92). Nykyään siirron onnistuminen on harvinaista (Allen, Tedi ja Shakeel 2014). Hostilla siirtoa tehdessä host-komentoon lisätään ”-l verkkotunnus nimipalvelin”. Mikäli DNS-vyöhykkeen siirto epäonnistuu, voidaan käyttää muita DNS-palvelimen kuulusteluun käytettäviä työkaluja kuten fierceä, joka tuottaa vaihtoehtoisia DNS-vyöhykkeen siirtokohteita (Engebretson ja Kennedy 2013, s.43). DNSEnum on työkalu, jolla voidaan suorittaa vyöhykkeen siirto fiercen tapaan (Baloch 2015, s.80).

Hosti, dig ja nslookup voivat suorittaa DNS-vyöhykkeen siirron ja ovat toiminnaltaan hyvin samanlaisia. Nslookupin ja digin toimintaa selvitettiin aiemmin luvussa 3.1. Passiivisesta tiedonkeruusta eroten, työkaluille annetaan kohteen DNS-palvelimen IP-osoite (Engebretson ja Kennedy 2013). Fierce ja DNSEnum eroavat edellä mainituista työkaluista siten, että ne yrittävät DNS-vyöhykkeen siirtoa. Mikäli vyöhykkeen siirto epäonnistuu, fierce käyttää raakaa voimaa (engl. brute-force) lähettämällä lukuisia kyselyitä kohteen DNS-palvelimelle (Engebretson ja Kennedy 2013, s.43). DNSEnum käyttää myös raakaa voimaa (Allen, Tedi ja Shakeel 2014). Raa’alla voimalla fierce yrittää saada haltuunsa isäntänimiä. Molemmat DNSEnum ja fierce ottavat verkkotunnuksen ja yrittävät tehdä vyöhykkeen siirron, jonka jälkeen molemmat käyttävät raakaa voimaa DNS-kuulusteluun (Allen, Tedi ja Shakeel 2014, 94–99).

Aktiiviseen tiedonkeruuseen kuuluu myös tiedon kerääminen sähköpostipalvelimelta. Tiedon keräämineen tapahtuu lähettämällä palvelimelle sähköposti, joka sisältää epäilyttävän tiedoston. Tavoitteena on saada hylkäysviesti, joka voi sisältää tietoja sähköpostipalvelimista. (Engebretson ja Kennedy 2013, s.44)

Passiivisen tiedustelun kappaleessa mainittiin muutama tiedonkeruun monitoimityökalu. Aktiivisen tiedonkeruun monitoimityökalu on WhatWeb. WhatWeb suorittaa aktiivisen tiedon-

keruun kohteen verkkosivuilta (Baloch 2015, s.62). Lähteen mukaan WhatWeb tunnistaa palvelinversion, kerää sähköpostiosoitteita ja antaa SQL-palautuskoodeja.

3.3 Google-hakkerointi

Google-hakkerointi on osa passiivista tiedonkeruuta, mutta laajuutensa vuoksi se käsitellään tässä tutkielmassa erillisenä kokonaisuutenaan. Google-hakkerointi on siis tiedonkeruuta Googlen tietokannasta. Googlen käytettävyys on laaja ja se on varmasti yksi syy Google-hakkeroinnin syntyyn. Tietokannat yleisesti saattavat pitää sisällään tietoa, joka on joskus julkaistu ja sittemmin poistettu (Shimonski, Bishop ja Zenir 2014, s.55). Lukuisissa käsitellyistä kirjallisuuslähteistä mainitaan Google-hakkerointi yhtenä tiedonkeruutapana, joten sitä voidaan pitää merkittävänä osana tiedonkeruuta.

Google-hakkeroinnin tarkoituksena on tehdä sellaisia kyselyitä, jotka tarjoavat tarkempia hakutuloksia (Long ym. 2008, s.13). Pyritään saamaan omia tarpeita vastaava tieto tietokannasta. Kun tuloksia on vähemmän, niiden läpikäynti on helpompaa. Google on hyvä työkalu kohdennettuihin hakuihin.

Aiemmin käsitellyistä tiedonkeruumenetelmistä käy ilmi, että verkkotunnus on tärkeä osa tiedonkeruuta. Näin on myös Google-hakkeroinnissa. Kirjassaan Long ym. (2008) esittelee monia eri operaattoreita Google kyselyiden tekemiseen. Tässä tutkielmassa käymme läpi sellaiset operaattorit, jotka esiintyvät myös muissa lähteissä, jotta osio ei ole turhan laaja. Operaattorit ovat sanoja, jotka rajaavat hakua. Kyselyitä luodessa pitää kiinnittää huomiota kielen oikeaan syntaksiin, jotta kyselyn operaattorista ei tule hakutermiä (Long ym. 2008, s.51). Hakuterminä operaattori ei rajaa hakuja, vaan tuottaa tuloksia, joissa operaattori on hakutermi.

Google-hakkerointi tehdään muodostamalla kysely ja syöttämällä se Googlen hakukoneeseen. Edellisessä kappaleessa käsiteltiin operaattoreita, joihin tarkennutaan nyt. Ensimmäinen käsiteltävä operaattori on ”site”, jolle syötetään verkkotunnus ilman välilyöntiä (Engebretson ja Kennedy 2013, s.26). Verkkotunnuksen jälkeen syötetään haettava termi tai termit. Esimerkki tällaisesta hausta olisi *site:verkkotunnus hakusanat*. Tämä haku palauttaa verkkosivulta tulokset, jotka sisältävät kyselyssä määritellyt hakusanat (Engebretson ja

Kennedy 2013, s.26). Site-operaattoria kannattaa käyttää, kun halutaan tietyltä verkkosivulta tietoa. Muita hyödyllisiä operaattoreita ovat ”intitle” ja ”allintitle”. Engebretson ja Kennedy (2013) kertovat, että edellä mainitut operaattorit palauttavat sellaisia sivuja, joiden otsikko sisältää joko kaikki tai osan hakusanoista. URLstä haku tapahtuu ”inurl”-operaattorilla, joka palauttaa vain ne URLt, jotka sisältävät halutun hakusanan (Chauhan ja Nutan 2015, s.79). Chauhan ja Nutan (2015) toteavat, että hauissa voidaan myös käyttää perinteisiä boolean operaattoreita kuten AND ja OR.

Välimuistista etsiminen voi antaa hyödyllistä tieto kohteesta. Välimuisti saattaa sisältää sellaista tietoa, joka on jo poistettu verkkosivulta. Välimuisti kysely tehdään käyttämällä ”cache”-operaattoria. Operaattorin jälkeen hakuun sijoitetaan verkkotunnus. Tiedostojen etsinnässä ”filetype” palauttaa vain sellaisia tiedostoja, mitkä haussa on määritelty. Voidaan hakea esimerkiksi vain PDF-tiedostoja. Operaattoreita voidaan yhdistää. (Engebretson ja Kennedy 2013, s.27-28)

Taulukko 1 sisältää esimerkkihakuja. Ensimmäinen haku palauttaa kaikki jyu.fi-sivut, jotka sisältävät sanan ”tietotekniikka”. Toinen haku palauttaa kaikki sivut, joiden otsikko sisältää sanan ”kissa”. Viimeinen haku palauttaa kaikki sivut, joiden URL sisältää sanan ”jyu”.

Taulukko 1. Esimerkkihakuja

operaattori:verkkotunnus/hakusana	hakusana
site:jyu.fi	tietotekniikka
intitle:kissa	-
inurl:jyu	-

3.4 Käyttäjän manipulointi

Social engineering eli käyttäjän manipulointi on Wilhelmin mukaan todella tehokas tapa kerätä tietoa, mutta siihen oltava lupa (Wilhelm 2013). Käyttäjän manipulointiin tarvitaan tietoja kohteesta, jotta manipulointi osataan kohdentaa. Käyttäjän manipulointi voidaan toteuttaa usealla eri tavalla ja hyökkäys voidaan jakaa ihmis- ja tietokonepohjaiseen manipulointiin (Patel 2013, s.7-8). Patelin (2013) mukaan ihmispohjainen hyökkäys voidaan toteuttaa hui-

jaamalla kohteen henkilöstöä, esiintymällä henkilökuntana tai käyttäjänä, salakuuntelemalla tai jopa kaivelemalla kohteen roskiksia (Patel 2013). Patel (2013) luettelee tietokonepohjaisiin hyökkäyksiin kuuluvan ponnahdusikkunat ja tietojen kalastelu (engl. phising). Esimerkiksi tietojenkalastelulla voidaan kalastella salasanoja ja muita henkilökohtaisia tietoja. SET on käyttäjän manipulointiin tarkoitettu työkalu (Patel 2013, s.10).

Hyökkäys, jossa käyttäjää manipuloidaan, on helppo toteuttaa kalastelemalla tietoja sähköpostitse. Hyökkäyksen onnistumiseksi pitää ensin kerätä tietoja kohteesta, jotta lähetettävä sähköposti vaikuttaa mahdollisimman luotettavalta. Kun sähköposti tai sen sisältämät linkit avataan, voidaan tietokoneeseen asentaa kuunteluohjelmia ja päästä käsiksi dataan. (Nelson ym. 2016)

Ihmispohjaisen hyökkäyksen voi toteuttaa esiintymällä toisena henkilönä. Penetraatiotestaaja voi esimerkiksi esiintyä kohteen työntekijänä. Testaaja voi salakuunnella kohteen henkilöstöä tai jopa yrittää päästä käsiksi yksityisviesteihin. Käänteinen käyttäjän manipulointi (engl. reverse social engineering) tarkoittaa sitä, että hyökkääjä yrittää houkutella uhrin itse ottamaan yhteyttä ja luovuttamaan tietoja. Testaaja voi esiintyä käyttäjänä ja yrittää kalastella salasanoja tätä kautta. (Patel 2013, s.8)

Aiemmin mainittu SET-työkalu on valmiiksi asennettu Kali Linuxiin. SETin avulla voidaan tehdä klooni verkkosivusta ja käyttää tätä kloonia käyttäjän huijaamiseen. SETin valikosta voi valita haluamansa hyökkäyksen kuten massasähköpostihyökkäyksen. Massasähköpostihyökkäyksessä kohteen sähköposti pyritään täyttämään turhilla viesteillä, jotta se saadaan kaatumaan. Lisäksi SETin käyttäjän manipulointi osiosta löytyy kahdeksan muuta hyökkäystyökalua. (Patel 2013, s.10-13)

Käyttäjän manipulointiin ei aina tarvita varsinaisesti käyttäjän manipulointiin tehtyjä työkaluja, vaan se voidaan toteuttaa esimerkiksi sosiaalisessa mediassa. Käyttäjän manipulointiin käytettäviä tapoja on useita, mutta niiden käytön kanssa pitää olla varovainen.

4 Tiedonkeruutapojen vertailu

Tiedonkeruun tapoja on monenlaisia, kuten luvussa 3 ja kuviossa 2 käsiteltiin. Erilaisia työkaluja on monenlaisia ja osa niistä muistuttaa hyvin paljon toisiaan tai tekee täysin saman asian, mutta eroaa käytettävyydeltään. Lisäksi sama tieto voidaan saada useasta eri lähteestä ja usealla eri tavalla. Jokainen tiedonkeruun yritys ei aina onnistu. Aktiivisessa tiedonkeruussa saatu tieto on voitu saavuttaa jo passiivisessa vaiheessa. Kuten aiemmin on todettu, aktiivisen tiedonkeruun kautta tieto voi olla paikkansapitävämpää eli päivitettyä. Tässä osioissa pohditaan tiedonkeruun tapojen näkyvyyttä ja näkyvyyden merkitystä. Lisäksi käsitellään tiedonkeruu tapojen ja työkalujen käytettävyyttä sekä eroavaisuuksia.

4.1 Näkyvyys

Penetraatiotestausta määritellessä todettiin, että penetraatiotestaus on luvanvaraista, joten voidaan pohtia, onko näkyvyydellä edes merkitystä. Tiedonkeruutapojen näkyvyyttä käsitellään lähteissä jakamalla tiedonkeruu aktiiviseen sekä passiiviseen. Jo tämä jako osoittaa, että näkyvyydellä on jonkinlaista merkitystä tiedonkeruussa. Allen, Tedi ja Shakeel (2014) toteavat kirjassaan, että kummassakin tavassa on omat hyötynsä ja passiivisessa tiedonkeruussa näihin hyötyihin lukeutuu huomaamattomuus. Broad, Bindner ja Bindner (2013) kertovat, että testauksessa pyritään imitoimaan vihamielistä hyökkääjää. Vihamielinen hyökkääjä todennäköisesti pyrkii minimoimaan näkyvyytensä. Henry (2012) painottaa aktiivisen tiedonkeruun vastatoimiksi systeemien aktiivista monitorointia ja lokitietojen tarkastelua. Vastatoimien toimivuutta voidaan testata suorittamalla aktiivista tiedonkeruuta. Käytännössä tämä tarkoittaa sitä, että tutkitaan, havaitseeko kohde hyökkääjän läsnäoloa. Pohdittaessa, onko penetraatiotestaajalla tarvetta piilotella jälkiään, voidaan vedota testaajan pyrkimykseen jäljitellä pahantahtoisen hyökkääjän käytöstä sekä kohteen tarpeeseen havaita mahdollinen hyökkäys. Testaajan kannattaa yrittää pysyä mahdollisimman huomaamattomana, koska näin taitava hyökkääjä tekisi ja kohteen on hyvä testata, havaitseeko tietoturva testaajan läsnäoloa. Näiden toteamuksen perusteella näkyvyydellä on merkitystä ja näkyvämpiä tiedonkeruutapoja tulisi sisällyttää testaukseen sen ollessa mahdollista.

Yksi harvoista passiivisen tiedonkeruun tavoista, joka voi jättää jäljen penetraatiotestaajasta kohteelle, on verkkosivun kopiointi. Engebretson ja Kennedy (2013) toteavat kirjassaan, että on otettava huomioon sivuston kopioinnin jäljitettävyys. Kopiointi voi jättää tietoja kohteen lokiin, mutta niitten huomaaminen riippuu lokitietojen seuraamisesta. “HTTrack Website Copier - Offline Browser” (2007) sivuston mukaan muillakin kuin penetraatiotestaajilla voi olla tarvetta kopioida verkkosivustoja tai osia niistä. Tästä voitaisiin päätellä, että lokitiedoista voi olla vaikeaa erottaa sivuston pahantahtoista ja hyväntahtoista kopioimista. Kopiointia on myös mahdollista rajoittaa eri tavoin (“HTTrack Website Copier - Offline Browser” 2007).

Aktiivisessa tiedonkeruussa taas siirrytään näkyvämmin kommunikoimaan kohteen kanssa luvussa 3.2 käsitellyillä tavoilla. Kohteen kanssa kommunikointi saattaa herättää huomiota ja tehdä penetraatiotestaajan läsnäolon tiettäväksi. Tämä ei välttämättä ole paha asia. Tiedonkeruun havaitseminen voi kertoa kohteen hyvästä suojauksesta. Kuten aiemmin on mainittu, osa aktiivisen vaiheen tiedosta on saavutettu jo passiivisessa tiedonkeruussa. Tämän takia voidaan pohtia, onko aktiivinen tiedonkeruu tarpeellista. Yleensä kuitenkin kohteen DNS-palvelin sisältää päivitetyn tiedon, kuten luvussa 3.2 todettiin. Tästä voidaan tehdä päätelmä, että aktiivinen tiedonkeruu on tärkeää, vaikka se onkin näkyvämpää kuin passiivinen, ja tiedot on saatettu saavuttaa jo passiivisin menetelmin.

Passiivisen tiedonkeruun osiossa käsiteltiin monitoimityökaluja kuten DMitry ja Maltego. Näistä Maltego käyttää tiedonkeruussa julkista tietoa (Allen, Tedi ja Shakeel 2014, s.102). Näkyvyyden kannalta Maltego on parempi työkalu, mikäli testaaja haluaa pysyä piilossa. DMirty menee jo porttien skannaukseen, joten sen käyttö on paljon näkyvämpää.

Sähköpostipalvelimen kanssa kommunikointi on myös hyvin näkyvää. Lisäksi käyttäjän manipulointi on näkyvää ja voidaan helposti yhdistää penetraatiotestaajaan. Käyttäjän manipulointi testaa kohteen henkilöstön tietoturvatietoisuutta ja kykyä tunnistaa käyttäjän manipulointihyökkäys. Sillanpää ja Hautamäki (2020) teettivät kyselyn testatakseen kohteen henkilöstön toimintaa ja tietoisuutta käyttäjän manipulointihyökkäyksessä.

Vaikka passiivinen tiedonkeruu on vaivihkaisempaa ja ei yhtä näkyvää kuin aktiivinen, on tärkeää sisällyttää penetraatiotestaukseen myös aktiivista tiedonkeruuta (Engebretson ja Ken-

nedy 2013). Baloch (2015) ei taas aina suosittele aktiivisen tiedonkeruun käyttöä sen näkyvyyden vuoksi. Täytyy muistaa, että aktiivinen tiedustelu luvatta voi aiheuttaa testaajalle ongelmia. Aktiivinen tiedonkeruu voi tarjota paljon tietoa esimerkiksi DNS-vyöhykkeen siirrolla, mutta kaikkia tietoja sillä ei voi saavuttaa. Esimerkiksi avainhenkilöistä ei saada välttämättä tietoa aktiivisella tiedonkeruulla, jollei käytetä käyttäjän manipulointia. Onkin suositeltavaa käyttää sekä näkyvää, että vaivihkaisempaa tiedustelua penetraatiotestauksessa. Penetraatiotestaajan kannattaa aina käydä läpi asiakkaan kanssa, mitkä tavat kuuluvat testaukseen (Allen, Tedi ja Shakeel 2014, s.85).

4.2 Käytettävyys

Tässä osiossa käsittelemme tiedonkeruutapojen käytettävyyttä. Wilhelm (2013) kirjassaan toteaa, että aktiivista tiedonkeruuta pidetään yleensä hyödyllisempänä kuin passiivista. Hän on kuitenkin itse eri mieltä tästä ja hän painottaa, että luottamuksellinen tieto on voinut vuotaa penetraatiotestaajan saataville. Passiivinen tiedonkeruu on siis käyttävyydeltään hyvä, koska sillä voidaan saada haltuun luottamuksellista tietoa. Engebretson ja Kennedy (2013) toteavat, että hyvä tiedustelu tarvitsee sekä aktiivista että passiivista tiedonkeruuta ollakseen kattava. Vaikka esimerkiksi aktiivisen tiedustelun DNS-vyöhykkeen siirto epäonnistuu usein, on se onnistuessaan valtava tietoturvariski ja tärkeä havaita. Vaikka siirto yleensä epäonnistuu, voi esimerkiksi varmuuskopiopalvelin sallia DNS-vyöhykkeen siirron (Baloch 2015, s.79).

Passiivisen tiedustelun osioissa käsiteltyä Archive.org-sivustoa ei kaikissa lähteissä mainita, vaikka se on erittäin hyvä työkalu kohteen verkkosivuja tutkittaessa. Sivusto on helppokäyttöinen ja vain muutamalla painaluksella pääsee tutkimaan kohteen verkkosivun historiaa. Archive.org-työkalun kanssa kannattaa kuitenkin olla varovainen, ettei vahingossa yhdistä kohteen verkkopalvelimeen (Wilhelm 2013, s.158). Mikäli kohteen verkkosivujen tietoja halutaan tutkia, Archive.org on hyvä työkalu siihen. Archive.org voi tarjota edellisessä kappaleessa mainittuja vuotaneita tietoja.

Passiivisiin tapoihin kuuluvat myös hakukoneet. Useassa lähteessä Google-hakkerointi nousee esille, joten sitä voidaan pitää käytettävänä tiedonkeruuntapana. Engebretson ja Kennedy

(2013) toteavat operaattorien käytön vaikuttavan merkittävästi hakutuloksiin. Kirjassaan Long ym. (2008) toteaa, että Google on yksinkertainen ja tehokas. Näistä syistä Google-hakkerointi kannattaa sisällyttää tiedonkeruuseen. Vaikka Google on kattava hakukone, tietojenkeruuseen olisi hyvä sisällyttää myös muita hakukoneita. Muitakin hakukoneita on hyvä käyttää, koska testauksen halutaan olevan perusteellinen (Engebretson ja Kennedy 2013, s.30). Luvussa 3.1 käsitelty Foca hakee metatietoa Googlen lisäksi muista hakukoneista.

Edellisessä kappaleessa käsitelty Foca on tiedostojen analysointiin käytetty työkalu (Baloch 2015). Focan lisäksi Metagoofil kerää myös metatietoa. Näiden työkalujen väliltä eroa on vaikea tehdä, koska esimerkiksi Engebretson ja Kennedy (2013) toteavat kirjassaan, että Metagoofil on loistava työkalu metatiedon keruuseen, mutta mainitsevat myös Focan. Baloch (2015) puolestaan toteaa Focan olevan tehokas työkalu tiedostojen analysointiin.

Verkkosivun kopioimiseen käytettävissä työkaluissa on eroja. Työkalut Wget ja HTTrack ovat luvussa 3.1 käsiteltyjä verkkosivun kopioimiseen tarkoitettuja työkaluja. Baloch (2015) toteaa, että työkaluista kattavin on HTTrack. Aloittelijalle HTTrack saattaa olla käyttäjäystävällisempi, koska se sisältää käyttöliittymän. Suuri osa penetraatiotestauksesta on kuitenkin komentorivillä työskentelyä, joten testauksen näkökulmasta käyttöliittymällä ei ole välttämättä väliä. Engebretson ja Kennedy (2013) eivät tiedonkeruussa edes mainitse Wgettiä ja Baloch (2015) toteaa, että HTTrack tarjoaa mahdollisuuden tutkia sivua syvällisemmin. Näiden lähteiden mukaan HTTrack on käytettävyydeltään parempi kuin Wget.

Aktiiviseen tiedonkeruuseen siirryttäessä käyttäjän manipulointia voidaan pitää erittäin tehokkaana tapana saada tietoa. Tämän toteaa Wilhelm (2013) kirjassaan sanomalla, että käyttäjän manipuloinnilla voidaan saada paljon enemmän tietoa kuin muilla tiedonkeruun tavoilla. Toisin sanoen käyttäjän manipulointia voidaan pitää luvallisesti tehtynä tehokkaana tiedustelun tapana. Tavan käytettävyyttä huonontaa ongelmat luvanvaraisuuden kanssa. Patel (2013) esittelee yksityisviestien lukemisen yhtenä tiedonkeruun metodina, mutta Suomen rikoslaissa tämä luokitellaan rikokseksi (“Rikoslaki 1889/39” 2021). Rikoslaissa myös todetaan, että toisena henkilönä esiintymistä voidaan pitää identiteettivarkautena. Käyttäjän manipuloinnissa on oltava tarkkana, missä rajoissa tätä tiedonkeruutapaa voidaan toteuttaa rikkomatta lakia tai ihmisen yksityisyyttä. Sillanpää ja Hautamäki (2020) tutkimuksessaan teettivät kyselyn yrityksen henkilöstölle. Kyselyssä yritettiin selvittää työntekijöiden toimintaa

käyttäjänmanipulointihyökkäyksessä. Tutkimuksessa pyrittiin myös saamaan käsitys työntekijöiden tietoturvatietoisuudesta. Käyttäjän manipuloinnin ongelmien vuoksi tällaisen kyselyn toteuttaminen voisi olla turvallisempaa kuin käyttäjänmanipulointihyökkäyksen varsinainen toteutus. Tuntemattomien muistitikkujen ympäriinsä jättäminen voisi myös olla turvallisempi tapa testata työntekijöitä. Muistitikut eivät kuitenkaan saisi sisältää todellisesti haitallisia ohjelmia. Luvallinen tietojenkalastelu voi olla varteenotettava tapa testata kohteen inhimillisen tekijän toimivuutta.

DNS-vyöhykkeen siirto tapahtuu hostilla, digillä sekä nslookupilla. Siirto antaa paljon tietoa ja on hyödyllinen tiedonkeruussa. Ongelmana siirrossa kuitenkin on, että yleensä se epäonnistuu. Aiemmin todettiin, että fierce käyttää siirron epäonnistuessa raakaa voimaa, jonka takia esimerkiksi Engebretson ja Kennedy (2013) suosittelevat sen käyttöä siirron epäonnistuessa. Heidän mukaansa fierce on myös helppokäyttöinen työkalu. Työkaluista nslookup ja dig ovat ominaisuuksiltaan samanlaisia, mutta dig sisältää kuitenkin enemmän toiminnallisuutta (Baloch 2015, s.77). Nslookup on näistä ensimmäinen ja siksi dig onkin voinut parantaa toimintaansa edeltäjäänsä verrattuna. Allen, Tedi ja Shakeel (2014) toteavat kirjassaan digin tulosteen olevan hostia selvempi ja digin olevan joustavampi käytettävyydeltään.

Passiivisessa sekä aktiivisessa tiedustelussa käsiteltiin muutamaa monitoimityökalua, kuten DMitry, Maltego sekä WhatWeb. Allen, Tedi ja Shakeel (2014) toteavat, että DMitry on erittäin kätevä työkalu, koska se suorittaa monta tiedonkeruuta ja tallentaa kaikki tiedot yhteen kansioon. Maltegon käyttöliittymä on selkeä ja kerätty tieto esitetään graafisesti ottaen huomioon tietojen väliset yhteydet (Allen, Tedi ja Shakeel 2014, s.102). Tieto kerätään yhtä työkalua käyttäen ja esitetään johdonmukaisesti. Nämä työkalut ovat siis varteenotettavia vaihtoehtoja tiedonkeruuseen helppoutensa ja johdonmukaisuutensa vuoksi.

Tiedonkeruun työkaluja on lukuisia, mutta tutkielmassa pääpaino oli eniten käsitellyillä työkaluilla. DNS-vyöhykkeen siirrossa suosituimmaksi työkaluksi nousi dig ja verkkosivun kopioimisessa HTTrack koettiin kaikista parhaimmaksi työkaluksi. Osalla työkaluista ja metodeista on hyvin vahva asema tiedonkeruussa. Vahva asema on pääteltävissä työkalujen esiintyvyydestä eri lähteissä. Tällaisina työkaluina voitaisi pitää whoisiä ja theharvesteria. Ongelmat käyttäjän manipuloinnissa huonontavat sen asemaa tällaisissa vertailuissa, kun taas Google-hakkeroinnin käytettävyyttä voidaan pitää suhteellisen vakaana.

5 Yhteenveto

Tutkielmassa käsiteltiin tiedonkeruuta ja siihen käytettäviä työkaluja. Tutkielmassa pyrittiin ottamaan selvää erilaisista tiedonkeruun tavoista sekä niiden eroavaisuuksista. Tutkielmassa selvisi, että työkaluja on monenlaisia eri tietojen keräämiseen. Osa työkaluista on myös tehty täysin saman tiedon keräämiseen. Työkalut saattavat saavuttaa saman tiedon, mutta erota käytettävyydeltään sekä näkyvyydeltään. Näkyvyydeltään aktiiviset tiedonkeruutavat olivat näkyvämpiä kuin passiiviset. Joissain tapauksissa aktiiviset tavat antoivat kuitenkin enemmän tietoa kuin passiiviset. Työkalujen käytettävyydessä oli eroja, mutta ne olivat suurelta osin pieniä lukuun ottamatta käyttäjän manipulointia, jonka käytettävyyttä voidaan tutkielman perusteella pitää heikkona. Käyttäjän manipulointia pitäisi käsitellä Suomen lainsäädännön puitteissa, jotta sen käytettävyys voisi olla parempi. Osalla työkaluista vaikuttaa olevan vahva asema, sillä ne esiintyvät useassa lähteessä, joita tutkielma käsittelee. Päätelmän paikkansapitävyyttä ei voida kuitenkaan pelkästään lähteiden perusteella todeta täysin varmaksi. Työkalujen käytettävyyteen testauksessa vaikuttaa testauksen laajuus, testaajan mieltymykset sekä tiedon saatavuus. Mikäli esimerkiksi käyttäjän manipulointia ei saa testauksessa suorittaa, on se työkaluineen jätettävä testauksesta. Tiedon saamiseen voidaan joutua käyttämään vaihtelevia työkaluja ja testaaja käyttää myös niitä työkaluja, jotka kokee itselleen mieleisiksi. Työkalujen käytettävyyden eroista saisi mahdollisesti paremman käsityksen suorittamalla tutkimuskyselyn penetraatiotestaajille.

Lähteet

Allen, Lee, Heriyanto Tedi ja Ali Shakeel. 2014. *Assuring Security by Penetration Testing: Master the Art of Penetration Testing with Kali Linux.* Packt Pub.

Baloch, Rafay. 2015. *Ethical Hacking and Penetration Testing Guide.* Boca Raton: CRC Press.

Broad, James, Andrew Bindner ja Andrew Bindner. 2013. *Hacking with Kali.* Syngress.

Chauhan, Sudhanshu, ja Kumar Panda Nutan. 2015. *Hacking Web Intelligence: Open Source Intelligence and Web Reconnaissance : Concepts and Techniques.* Syngress.

Engebretson, Patrick Henry, ja David Kennedy. 2013. *The Basics of Hacking and Penetration Testing: Ethical Hacking and Penetration Testing Made Easy.* Syngress.

Henry, Kevin M. 2012. *Penetration Testing: Protecting Networks and Systems.* Ely, Cambridgeshire, U.K: IT Governance Pub.

"HTTrack Website Copier - Offline Browser". 2007. Viitattu 15. maaliskuuta 2022. https://www.httrack.com/html/abuse.html.

"Kali". 2022. Viitattu 5. helmikuuta 2022. https://www.kali.org/.

Long, Johnny, Justin Brown, Bill Gardner, Bill Gardner ja Justin Brown. 2008. *Google Hacking for Penetration Testers Volume 2.* Syngress.

Nelson, Jennifer, X. Lin, C. Chen, J. Iglesias ja J. J. Li. 2016. "Social Engineering for Security Attacks". (New York, NY, USA), https://doi.org/10.1145/2955129.2955158.

Patel, Rahul Singh. 2013. *Kali Linux Social Engineering.* Birmingham: Packt Publishing.

"Rikoslaki 1889/39". 2021. Viitattu 8. maaliskuuta 2022. https://www.finlex.fi/fi/laki/ajantasa/1889/18890039001.

Shimonski, Robert, Allison Bishop ja John Zenir. 2014. *Cyber Reconnaissance, Surveillance and Defense.* Burlington: Syngress.

Sillanpää, Miika, ja Jari Hautamäki. 2020. “Social Engineering Intrusion: A Case Study”. (New York, NY, USA), https://doi.org/10.1145/3406601.3406631.

Wilhelm, Thomas. 2013. *Professional Penetration Testing: Creating and Learning in a Hacking Lab.* Burlington: Elsevier Science.