



Detektion av illasinnad exfiltrering av data via nätverk

En systematisk litteraturgenomgång

Henrik Karlzén och Christian Valassi

Henrik Karlzén och Christian Valassi

Detektion av illasinnad exfiltrering av data via nätverk

En systematisk litteraturgenomgång

Titel	Detektion av illasinnad exfiltrering av data via nätverk – En systematisk litteraturgenomgång
Title	Detection of malicious exfiltration of data via networks – A systematic literature review
Rapportnr/Report no	FOI-R--5376--SE
Månad/Month	December
Utgivningsår/Year	2022
Antal sidor/Pages	69
ISSN	1650-1942
Uppdragsgivare/Client	Försvarsmakten
Forskningsområde	Informationssäkerhet
FoT-område	Operationer i cyberdomänen
Projektnr/Project no	E716587
Godkänd av/Approved by	Christian Jönsson
Ansvarig avdelning	Ledningssystem

Bild/Cover: iStockphoto

Detta verk är skyddat enligt lagen (1960:729) om upphovsrätt till litterära och konstnärliga verk, vilket bl.a. innebär att citering är tillåten i enlighet med vad som anges i 22 § i nämnd lag. För att använda verket på ett sätt som inte medges direkt av svensk lag krävs särskild överenskommelse.

This work is protected by the Swedish Act on Copyright in Literary and Artistic Works (1960:729). Citation is permitted in accordance with article 22 in said act. Any form of use that goes beyond what is permitted by Swedish copyright law, requires the written permission of FOI.

Sammanfattning

Den här rapporten presenterar en systematisk litteraturgenomgång om detektion av illasinnad nätverksbaserad dataexfiltrering. Sådan exfiltrering utgörs av angripares överföring av information från målmaskin till angriparmaskin. Eftersom det är mycket svårt att hålla alla angripare ute ur nätverken är detektionen av exfiltrering en viktig del i försvaret.

Litteraturgenomgången inkluderar 48 forskningsartiklar utgivna 2012–2022. Genomgången visar att artiklarna har tydliga syften men däremot oftast ottydligt beskrivna hypoteser. Därtill visar genomgången att artiklarna fokuserar på labbmiljöer som efterliknar universitetsnät medan bara en artikel har en militär nätverksmiljö. Artiklarna beskriver i undantagsfall vilka data exfiltreringen rör. Det rör sig då om kreditkortsuppgifter, inloggningsuppgifter och dokument. Det är också ovanligt att artiklarna beskriver tänkta hotaktörer.

Artiklarna försöker detektera exfiltrering som döljs på olika sätt. Oftast rör det sig om döljande i form av placering av data i pakethuvuden, användning av protokoll som normalt inte nyttjas för användares dataöverföring eller av kryptering. Artiklarnas exfiltrering genomförs oftast via protokollet DNS, vilket har sitt legitima bruk i översättning av domännamn till IP-adresser.

Detektionsmetoderna baseras på olika förändringar som exfiltreringen ger upphov till. Det rör sig om nätverksmässiga skillnader i entropi, tidsaspekter, stränglängder, trafikflöden samt pakethuvudinnehåll. Algoritmerna är vanligen baserade på djupinlärning eller traditionell maskininlärning. Detektionsmetoderna utvärderas relativt knapphändigt i artiklarna.

Rapportens bedömning är att mer forskning behövs om framförallt detektion av exfiltrering som sker via videomöten, blockkedjenätverk, DNS över HTTPS, IPv6 och andra nyare protokoll. Därtill behövs forskning som snarare än att försöka detektera specifika tekniker tar fram mer generella algoritmer som kan detektera fler exfiltreringstekniker.

Nyckelord: exfiltrering, dataexfiltrering, detektion, it-säkerhet, cyber, DNS, systematisk litteraturgenomgång

Summary

This report presents a systematic literature review of the detection of malicious network-based data exfiltration. Such exfiltration consists of attackers' transfer of data from target machine to attacker machine. Since it is not possible to keep all attackers out of networks, the detection of exfiltration is an important part of the defence.

The literature review includes 48 research papers that were published in 2012–2022. The review shows that the papers have clear aims, but, on the other hand, the hypotheses are mostly described vaguely. In addition, it is shown that the papers focus on lab environments that mimic university networks, while only one paper has a military network environment. The articles rarely describe the exfiltrated data, in the form of credit card details, login details and documents. It is also unusual for the articles to describe the supposed threat actors.

In exfiltration techniques, data is usually hidden by placing data in packet headers, using protocols that are not normally used for user data transmission, and by encrypting. The protocol usually used for exfiltration is DNS, which has its legitimate use in translating domain names to IP addresses.

Exfiltration is usually detected through network-based differences in entropy, time aspects, string lengths, traffic flows, and packet header content. The algorithms are usually based on deep learning, or traditional machine learning. The detection methods are evaluated relatively rarely in the papers.

The report's assessment is that more research is needed primarily on the detection of exfiltration that occurs via video conferencing, blockchain networks, DNS over HTTPS, IPv6, and other newer protocols. In addition, there is a need for research that rather than detecting specific techniques, takes a more general approach, with algorithms that can detect more exfiltration techniques.

Keywords: exfiltration, data exfiltration, detection, cyber, DNS, SLR, systematic literature review

Innehållsförteckning

1	Inledning	9
1.1	Syfte och mål	10
1.2	Avgränsning	10
1.3	Läsanvisning	10
2	Bakgrund	12
2.1	Ett exempel på dataexfiltrering	12
2.2	Dataexfiltreringens konsekvenser	13
3	Relaterat arbete	16
3.1	Relaterad FOI-forskning	16
3.2	Genomgångar av studier om dataexfiltrering och närliggande områden	17
4	Metod	20
4.1	Sökning efter artiklar	22
4.2	Kontroll av artiklarnas relevans	23
4.3	Extrahering av data	25
4.4	Syntes av data	27
5	Artiklarnas ursprung och kvalitet	28
5.1	Ursprung	28
5.2	Citeringar	29
5.3	Subjektiv kvalitetsbedömning	29
6	Miljön för exfiltreringen	31
6.1	Legitima datas ursprung	31
6.2	Nätverken	32
7	Angreppsscenario för exfiltreringen	34
7.1	Data som exfiltreras	34
7.2	Hotaktörerna	34
8	Dataexfiltreringsteknikerna som ska detekteras	36
8.1	Exfiltreringens döljande och protokoll	36
8.2	Verktyg och dataset	38
9	Metoder för att detektera illasinnad dataexfiltrering	40

9.1	Kännetecken som detektionen baseras på	40
9.2	Algoritmer för detektionen	42
9.3	Detektionsmetodernas effektivitet och nackdelar	45
10	Diskussion	48
10.1	Litteraturgenomgångens kvaliteter samt skillnader mot andra genomgångar	48
10.1.1	Litteraturgenomgångens styrkor och svagheter	48
10.1.2	Jämförelse med andra litteraturgenomgångar	49
10.2	Praktiska implikationer	50
10.2.1	Militära miljöer	50
10.2.2	Riktiga angrepp och exfiltreringsverktyg	52
10.2.3	Produkter för detektion av dataexfiltrering	53
10.2.4	Komplement till detektion	53
10.3	Framtida forskning som behövs	54
11	Slutsatser	56
11.1	Vilka metoder för detektion av illasinnad nätverksbaserad exfiltrering av data från it-system studeras i forskningslitteraturen?	56
11.1.1	Vad är studiernas ursprung och hur är deras kvalitet?	56
11.1.2	Vilken miljö antar studierna att exfiltreringen sker i? ..	56
11.1.3	Vilket angreppsscenario antar studierna att exfiltreringen sker i?	57
11.1.4	Vilka exfiltreringstekniker försöker studierna detektera?	57
11.1.5	Hur fungerar studiernas detektionsmetoder?	57
11.2	Rapportens bidrag	58
11.3	Framtida forskning	58
11.4	Avslutning	59
12	Referenser	60
12.1	Allmänna referenser	60
12.2	Relaterade FOI-rapporter	62
12.3	Genomgångar	63

12.4 Identifierade forskningsartiklar	64
---	----

1 Inledning

De detekterar aldrig dataexfiltrering...det är som om de är blinda.

It-säkerhetsexperten Kevin Mitnick om företags förmåga i samband med säkerhetstester (föreläsning hos it-säkerhetsföretaget Palo Alto Networks, 21 april 2022).¹

Bland it-säkerhetsexperter är det allmänt accepterat att det inte är möjligt att helt och hållet stoppa alla angripare från att ta sig in i försvarares it-system. En viktig del av försvaret av it-system är därför att hantera situationer där angripare redan tagit sig in. I sådana situationer krävs ett försvar på djupet som går ut på att inte låta angriparen göra skada inne i systemet och på att hantera angriparens utförsel (exfiltrering) av konfidentiella data. Denna rapport fokuserar på det senare, det vill säga dataexfiltrering.

Dataexfiltrering definieras som *otillåten överföring av information från ett system* (Nist, 2020) och utgör *överföringen av information från målmaskin till angriparmaskin* (Antwerp, 2011).

Angripare kan nyttja en stor mängd olika tekniker för att exfiltrera data. Dessa tekniker är ofta nätverksbaserade där exfiltreringen kan ske med allt ifrån enkla e-postmeddelanden till avancerade tekniker där data som ska exfiltreras läggs in som små förändringar i vanliga nätverkspaket. Exfiltreringsteknikerna kan också vara fysiska och exempelvis använda utskrifter och USB-minnen. Fysisk exfiltrering studeras dock inte i denna rapport.

Förutom att angripare har många exfiltreringstekniker till sitt förfogande är en viktig aspekt den ständiga ökningen av mängden värdefulla data som lagras i organisationers nätverk. Att värdena är större ökar sannolikheten att angripare vill exfiltrera data samtidigt som det blir större konsekvenser om exfiltreringen lyckas. Utmaningarna för försvararna är därmed stora. Ett sätt att ta sig an utmaningarna är att försöka detektera exfiltrering när den sker. Denna rapport beskriver en systematisk litteraturgenomgång av forskningsartiklar om detektion av illasinnad nätverksbaserad dataexfiltrering.

¹ Citaten i inledningen har översatts från engelska.

1.1 Syfte och mål

Rapporten har tagits fram i projektet *Analys av koncept och teknik för cyberförsvar och informationssäkerhet* inom ramen för Försvarsmaktens samlingsbeställning inom forskning och teknikutveckling (FoT). I projektet görs analyser av olika cybersäkerhetsproblem för att bland annat inrikta framtida forskningssatsningar.

Syftet med rapporten är att förbättra möjligheterna för de som vill försvara sina datornätverk genom att detektera illasinnad dataexfiltrering. Rapporten är ganska teknisk men är tänkt att kunna läsas av alla som har grundläggande kunskap om nätverkssäkerhet. Rapporten besvarar följande forskningsfråga:

1. *Vilka metoder för detektion av illasinnad nätverksbaserad exfiltrering av data från it-system studeras i forskningslitteraturen?*

Svaret på forskningsfrågan behöver förstås i en kontext. Den kontexten utgörs av vad det är för typ av forskning, med dess ursprung och kvalitet samt vilka miljöer, angreppsscenarioer och exfiltreringstekniker som forskningen om detektion studerar. Därför besvarar rapporten även följande forskningsfrågor:

2. *Vad är studiernas ursprung och hur är deras kvalitet?*
3. *Vilken miljö antar studierna att exfiltreringen sker i?*
4. *Vilket angreppsscenario antar studierna att exfiltreringen sker i?*
5. *Vilka exfiltreringstekniker försöker studierna detektera?*
6. *Hur fungerar studiernas detektionsmetoder?*

1.2 Avgränsning

Rapporten avgränsar sig till nätverksbaserad exfiltrering vid nätverkets gräns. Därmed inkluderas inte fysisk exfiltrering (via t.ex. USB-minnen) eller ren prevention längre in i ett nätverk (vid förberedelser till exfiltrering). Inte heller inkluderas senare detektion längre från nätverket (till exempel vid användning av exfiltrerade data). Därtill ska detektionen ske i realtid snarare än genom att i efterhand utföra forensiska utredningar.

1.3 Läsanvisning

Kapitel 2 ger en bakgrund om hur dataexfiltrering kan utföras samt vilka konsekvenser exfiltreringen kan få. Denna bakgrund behöver inte läsas av läsare som redan känner till grunderna om dataexfiltrering.

Kapitel 3 presenterar tidigare relaterad forskning, dels från FOI, dels från genomgångar inom dataexfiltrering och närliggande forskningsområden. Denna relaterade forskning kan vägleda intresserade läsare till mer forskning utöver denna rapport.

Kapitel 4 redogör för metoden för den systematiska litteraturgenomgången av forskningsartiklar. Denna metodbeskrivning är främst lämpad att läsas av andra forskare.

Kapitel 5 beskriver artiklarnas ursprung, deras antal citeringar samt en subjektiv bedömning av deras kvalitet. Dessa beskrivningar inleder rapportens rapportkapitel vilka bör läsas av den som vill skapa sig förståelse för rapportens resultat och själv kunna dra slutsatser från dem. Övriga läsare kan se sådana resultatkapitel som torftiga och kan vara mer betjänta av att gå direkt till diskussionskapitlet.

Kapitel 6 beskriver vilka miljöer artiklarna använder i sina experiment om exfiltrering utifrån legitima datas ursprung samt typen och storleken på nätverken som används.

Kapitel 7 beskriver angreppsscenariot för experimenten utifrån typen av data som exfiltreras samt vilka hotaktörerna är tänkta att vara.

Kapitel 8 beskriver vilka exfiltreringstekniker för exfiltrering som artiklarna försöker detektera utifrån hur exfiltreringen döljs och vilka protokoll som används samt utifrån vilka verktyg och dataset som används för att utföra exfiltreringen.

Kapitel 9 beskriver detektionsmetoderna utifrån vilka indata (kännetecken) de använder, vilka algoritmer som används samt metodernas effektivitet och nackdelar.

Kapitel 10 diskuterar resultaten utifrån litteraturgenomgångens kvaliteter och dess skillnad mot andra genomgångar. Därtill diskuteras vilka praktiska implikationer som finns gentemot en militär kontext, riktiga angrepp, produkter för riktiga angrepps detektion samt vilka försvar det finns som kan komplettera detektion. Dessutom diskuteras vilken framtida forskning som behövs.

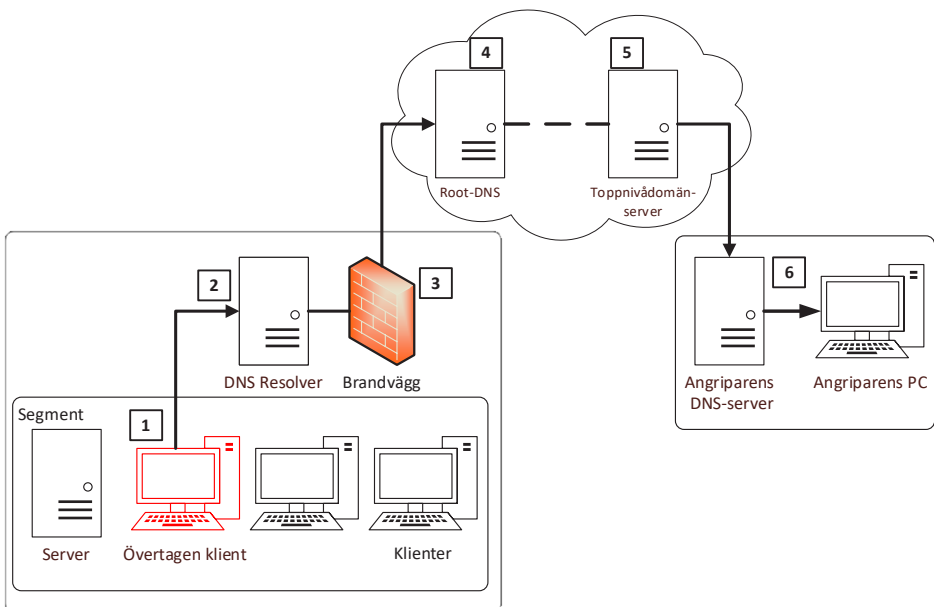
Kapitel 11 summerar slutsatser av resultaten och diskussionen i form av svaren på forskningsfrågorna, vad rapportens bidrag är gentemot annan forskning, vilken ytterligare forskning som behövs samt i allmänhet vad läsaren bör ta med sig från rapporten.

2 Bakgrund

I detta kapitel beskrivs ett exempel på nätverksbaserad exfiltrering samt vilka konsekvenser exfiltreringen kan ha.

2.1 Ett exempel på dataexfiltrering

I detta avsnitt ges ett exempel på dataexfiltrering. I exemplet har en angripare lyckats ta sig in i en organisations nätverk och samlat in data på en klient. Angriparen vill därefter exfiltrera dessa data från den övertagna klienten till angriparens nätverk. För att exfiltreringen inte ska detekteras lägger angriparen in data i DNS-paket vilka normalt inte används för överföring av användardata utan för att ställa förfrågningar om vilken IP-adress som tillhör vilket domännamn. Processen för exfiltreringen sker enligt Figur 1.



Figur 1: Dataexfiltrering med hjälp av DNS-förfrågningar.

I enlighet med figuren är processen följande:

1. Från den övertagna klienten skapar angriparen en DNS-förfrågan för ett domännamn där den första (och mest specifika) delen innehåller exfiltrerade data enligt: *EXFILTRERADEDATA.angriparensdomän.se*. Resterande del av domännamnet tillhör angriparens server. DNS-

förfrågan skickas till organisationens interna DNS-server (benämnd DNS Resolver).

2. Organisationens interna DNS-server vet inte vad IP-numret är för den efterfrågade domänen men antar att svaret finns hos en extern DNS-server utanför nätverket. DNS-förfrågan skickas därför till det interna nätverkets brandvägg som ligger på vägen ut ur nätverket.
3. Brandväggen ser att det rör sig om en DNS-förfrågan vilket brandväggen inte kan tolka som exfiltreringsdata. Brandväggen låter därför förfrågan passera vidare ut ur nätverket.
4. I det externa nätverket (internet) tar en extern DNS-server (benämnd Root DNS) emot paketet och försöker besvara förfrågan. Denna DNS-server vet dock inte heller svaret men känner till att för domäner som slutar på *.se* behöver förfrågan skickas vidare till en viss toppnivådomänserver och skickar därför förfrågan vidare.
5. Toppnivådomänservern kan inte heller svara på förfrågan men vet att för domäner som slutar på *angriparensdomän.se* behöver förfrågan skickas vidare till den DNS-server som äger domännamnet. Förfrågan skickas därför vidare till vad som är angriparens DNS-server.
6. Angriparens DNS-server tar emot DNS-förfrågan, förstår att det rör sig om exfiltrering, utvinner *EXFILTRERADEDATA* och skickar exfiltreringsdata vidare till angriparens dator. I normala fall hade det varit DNS-servers uppgift att besvara DNS-förfrågan och svaret hade då gått hela vägen tillbaka till den övertagna klienten. Vid exfiltrering är besvarandet inte lika relevant även om det kan ske ändå för att övriga DNS-servrar inte ska vänta på ett svar eller för att exfiltreringsverktyget ska få ett kvitto på att exfiltreringen fungerade.

Det finns många varianter på denna process. Mer om olika exfiltreringstekniker som kan användas i en dylik process beskrivs senare i rapporten.

2.2 Dataexfiltreringens konsekvenser

Det är svårt att bedöma konsekvenserna för en organisation som utsatts för dataexfiltrering. Detta beror bland annat på att det finns många olika typer av konsekvenser att beakta. Det är också möjligt att angripare förstör loggar och andra spår efter exfiltreringen eller att den sker så dolt att den inte alls detekteras. Därtill kan den som begått misstag som möjliggjort exfiltreringen vilja hemlighålla, förvränga eller förneka uppgifter om händelseförlopp och konsekvenser. För en bedömare utanför organisationen är det dessutom svårt att hitta objektiva källor som beskriver skeenden och kostnader på ett tillförlitligt sätt. Exempelvis slutar eventuella stämningar typiskt med förlikning där detaljer inte avslöjas för allmänheten.

I Tabell 1 sammanfattas några omtalade och stora exempel på dataexfiltrering. Som direkta konsekvenser räknas de som tydligt rör det nämnda offret som data exfiltreras från. Som indirekta konsekvenser räknas de som drabbar andra vilket kan få följder för organisationen, såsom förlorade kunder.

Tabell 1: Exempel på dataexfiltreringsincidenter med kostnader i miljoner kronor (mkr). Årtalet anger när exfiltreringen skedde, oavsett detektionsdatum. Allt rör illasinnad exfiltrering utom genomsnittet som även rör oavsiktliga dataläckage. Om ingen källa anges för konsekvenserna är källan densamma som i datakolumnen.

Offer (bransch)	År	Data (typ och mängd)	Konsekvenser
USA (myndigheter)	2010– 2012	Dokument på 740 000 konfidentiella sidor (vilka spriddes via Wikileaks) (Leopold, 2019).	Direkt: USA:s myndigheter uppgav att stödet för utländska operationer påverkades negativt. Indirekt: Säkerheten uppgavs äventyras för utländska samarbetspartner.
Marriott (hotell)	2014	Drygt 300 miljoner kunders personuppgifter (ICO, 2018).	Direkt: 230 mnkr i böter (ICO, 2018). Indirekt: Möjliggjorde enligt analytiker kartläggning av politikernas resvanor (Packetlabs, 2018).
British Airways (flyg)	2018	Drygt 400 000 kunders personuppgifter inklusive kreditkortsuppgifter (BBC, 2020).	Direkt: 250 mnkr i böter (BBC, 2020). Indirekt: Möjligen stulna pengar från kunder (BBC, 2018).
Mitsubishi (försvarsteknik med mera)	2020	Tidiga specifikationer för en hypersonisk robot (under anbudsprocess) (Sudo och Sato, 2020).	Direkt: Möjligen förlorad upphandling. Indirekt: Riskerade enligt uppgift delar av utvecklingen av japansk försvarsförmåga.
<i>Genomsnitt</i> (alla branscher)	2021– 2022	Normalstor mängd kunders personuppgifter (vare sig läckt illasinnat eller ej) (IBM, 2022).	Direkt: 44 mnkr (totalt, i genomsnitt). Indirekt: oklart.

På grund av den generella ovisshet som finns gällande beräkningar av kostnader bör dessa siffror endast ses som en uppskattning. Detta gäller inte minst då inget av de specifika fallen tar upp kostnader för interna utredningar, återställning och försämrat rykte.

3 Relaterat arbete

Detta kapitel introducerar tidigare genomförd forskning som är relevant för området dataexfiltrering. Kapitlet tar upp tidigare rapporter skrivna av FOI och beskriver dessa kortfattat. Därefter görs en genomgång av tidigare genomförda litteraturgenomgångar inom området dataexfiltrering och närliggande forskningsområden.

3.1 Relaterad FOI-forskning

Tidigare relaterad forskning som utförts av FOI resulterade i följande rapporter och memon:

- Lundholm m.fl. (2010) om dolda kanaler.
- Stenborg m.fl. (2010) om vattenmärkning.
- Karlzén och Sommestad (2022) om hotjakt.

Lundholm m.fl. (2010) ger läsaren en överblick av vad som innefattas i begreppet dolda kanaler och beskriver forskningsfältet relaterat till ämnet. Det beskrivs att dolda kanaler har till syfte att dölja kommunikation såsom exfiltrering av data. En rad olika typer av dolda kanaler beskrivs samt potentiella motmedel mot kanalerna. Därtill nämns att det finns ett stort antal dolda kanaler då i stort sett vad som helst relaterat till kommunikation kan utgöra en dold kanal så länge en avsändare kan påverka något som en mottagare kan observera. Rapporten beskriver att det behövs forskning kring hur problemet kring dolda kanaler ska hanteras när system som tidigare varit isolerade från varandra kopplas samman.

Stenborg m.fl. (2010) beskriver hur vattenmärkning kan användas på sensordata i syfte att skydda sensordata mot påverkan och läckage. Om exempelvis ett visst exemplar av en bild innehåller ett unikt vattenmärke kan betrodde användare avskräckas från att sprida bilden på felaktigt sätt eftersom en spridd kopia av bilden kan spåras tillbaka till läckan. Angripare kan försöka ge sig på vattenmärkningen i syfte att förstöra den. Sådana angrepp kan exempelvis utgöras av förstörande komprimering eller filtrering. Robust vattenmärkning bör därför kunna hantera dessa typer av angrepp.

Karlzén och Sommestad (2022) beskriver hotjakt och indikatorer på kompromettering (eng. indicators of compromise, IoC). Specifikt behandlar memot hur träning och övning ska utföras för att öka kompetensen inom hotjakt. I studien nyttjades ramverket Mitre ATT&CK och dess innehållande angreppstekniker för att utvärdera vilka IoC:er som är relevanta i träningsssammanhang med fokus på till exempel förbättring av förmågan att detektera exfiltrering.

3.2 Genomgångar av studier om dataexfiltrering och närliggande områden

Det finns sedan tidigare flera vetenskapliga genomgångar av forskningsartiklar inom dataexfiltrering och närliggande områden. Sådana genomgångar identifierades inledande i databasen Scopus. Därefter identifierades även genomgångar i den systematiska litteraturgenomgångens sökning vilken beskrivs i metodkapitlet. Genomgångar som bedömdes vara alltför långt från rapportens ämne exkluderades. De identifierade genomgångarna beskrivs i Tabell 2. Bland annat beskrivs huruvida genomgången varit systematisk (SLR) vilket den räknas som om den varit tydlig med hur dess artiklar identifierats, exempelvis genom att ange vilken söksträng och vilka databaser som söktes i.

Tabell 2. Tidigare genomgångar av forskningsartiklar inom dataexfiltration och närliggande områden. För varje genomgång nämns dess ämne, hur många (#) forskningsartiklar som ingick i genomgången, huruvida genomgången var systematisk (SLR), om genomgången publicerats i en tidskrift (T) eller i en akademisk bok (B), samt hur många andra forskningsartiklar som citerat genomgången enligt Google Scholar.

Artikel	Ämne	#	SLR	Publ	Cit
Shabtai m.fl. (2012)	Prevention av dataförlust (DLP)	37	Nej	B	225
Wendzel m.fl. (2015)	Nätverksbaserade dolda kanaler	109	Nej	T	146
Alneyadi m.fl. (2016)	Prevention av dataförlust (DLP)	25	Nej	T	134
Cheng m.fl. (2017)	Motmedel mot dataläckor	16	Nej	T	165
Ullah m.fl. (2018)	Dataexfiltreringstekniker och deras detektion	108	Ja	T	77
Conti m.fl. (2018)	Sidokanaler för mobila enheter	59	Nej	T	72
Akram m.fl. (2020)	Detektion av sidokanaler baserade på processorers cache	23	Nej	T	9
Tian m.fl. (2020)	Nätverksbaserade dolda kanaler	46	Nej	T	5
Avila m.fl. (2021)	Detektion av dataläckor med hjälp av säkerhetsloggar	33	Ja	T	5
Caviglione (2021)	Motmedel mot dolda kanaler	55	Nej	T	17
Wang m.fl. (2021)	Detektion av DNS-tunnlar (vilka bl.a. används för exfiltrering)	29	Nej	T	12
Sabir m.fl. (2022)	Maskininlärningsbaserad detektion av dataexfiltrering	92	Ja	T	6

Det kan noteras att en genomgång utgörs av en bok (Shabtai m.fl., 2012), tre har fokus på annat än detektion och andra motmedel (Wendzel m.fl., 2015; Conti m.fl., 2018; Tian m.fl., 2020) samt att en saknar fokus på nätverk (Akram m.fl., 2020). Övriga sju identifierades som en del av den systematiska litteraturgenomgången vilken beskrivs senare i denna rapport. Av dessa sju artiklar är det bara tre som själva är systematiska (Ullah m.fl., 2018; Avila m.fl., 2021; Sabir m.fl., 2022) trots att systematiska genomgångar krävs för att ge en

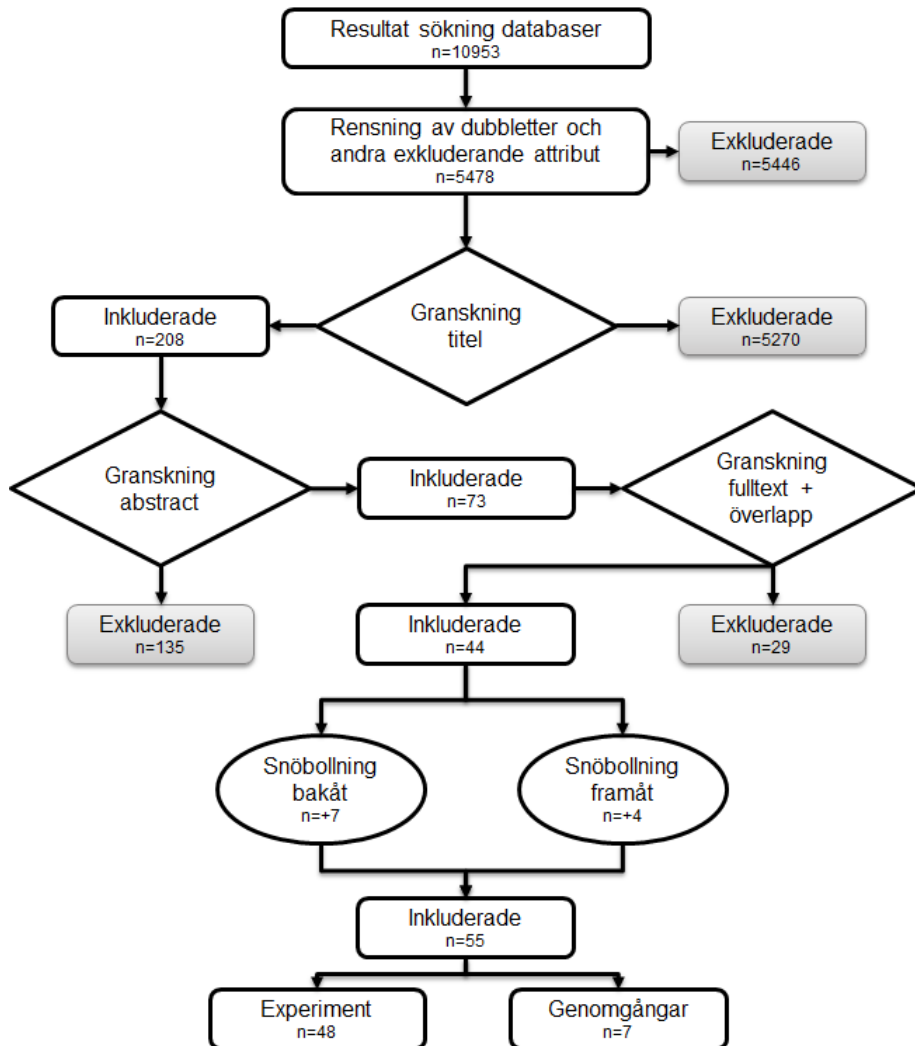
gedigen och rättvis bild av ett ämne och därmed bidra med ett större vetenskapligt värde (Kitchenham och Charters, 2007).

För de tre systematiska genomgångarna skiljer sig ämnesområdena åt en del. Ullah m.fl. (2018) rör dataexfiltration i allmänhet inklusive motmedel mot exfiltreringen. Avila m.fl. (2021) har fokus på att detektera dataläckor med hjälp av olika typer av loggar. Sabir m.fl. (2022) undersöker hur dataexfiltrering kan detekteras med hjälp av maskininlärning. Vår systematiska genomgång är ämnesmässigt närmast Ullah m.fl. (2018). Två skillnader är dock att vi har större fokus på detektion av exfiltrering samt inkluderar forskningsartiklar som publicerats efter Ullah m.fl. (2018).

4 Metod

Rapporten baseras i huvudsak på akademisk forskningslitteratur identifierad genom en så kallad systematisk litteraturgenomgång (eng. systematic literature review, SLR) utgående från Kitchenham och Charters (2007).

Metoden genomförs genom att: systematiskt söka efter artiklar med hjälp av en viss söksträng och i vissa databaser (se avsnitt 4.1); inkludera eller exkludera artiklar baserat på halvautomatisk och manuell applicering av ett antal kriterier (se avsnitt 4.2); extrahera data från de inkluderade artiklarna (se avsnitt 4.3); skapa en syntes av extraherade data för att måla upp en helhetsbild och dra gemensamma slutsatser från artiklarna (se avsnitt 4.4). En överblick över de två första delarna med sökning och inkludering ges av Figur 2.



Figur 2. Övergripande beskrivning av de centrala stegen i litteratursökningen och granskningsprocessen. Snöbollning bakåt avser artiklars referenser medan snöbollning framåt avser vilka andra som citerar artiklarna.

De olika stegen i figuren samt övriga delar av metoden beskrivs närmare nedan.

4.1 Sökning efter artiklar

Söksträngen togs fram genom att utgå från:

- Denna rapports forskningsfrågor.
- De söksträngar som använts i tidigare litteraturgenomgångar inom samma och snarlika områden. Sådana genomgångar listas i avsnitt 3.2.
- Hittills identifierade begrepps synonymer och andra närliggande begrepp.
- Explorativa sökningar för att se till att träffarna blev relevanta och till exempel inkluderade relevanta tidigare litteraturgenomgångar på området och andra relevanta artiklar som var kända sedan tidigare av rapportens författare.

Därtill begränsades sökningen till att enbart leta efter artiklar på engelska och som var färdiga och publicerade snarare än tidiga preliminära utgåvor. Dessutom begränsades sökningen i tid till artiklar publicerade från år 2012 och fram till dess att sökningen genomfördes tidigt under 2022.

Söksträngen applicerades i flera olika databaser som indexerar forskningsartiklar och tillhörande metadata. Vilka databaser som väljs spelar roll för resultatet och det är inte uppenbart hur databaserna bör väljas. Kitchenham och Brereton (2013) beskriver att på it-området täcks artiklar från de viktigaste tidskrifterna och konferenserna in av databaserna IEEE Xplore och ACM Guide to Computing Literature. Dessa två databaser valdes därför. För att täcka in ytterligare artiklar valdes även databaserna Scopus och Web of Science.

Sökningen gjordes 16 april 2022 i varje databas och då i artiklarnas metadatafält titel, sammanfattning (eng. abstract) och nyckelord. Exakt hur söksträngen utformades varierade beroende på de olika databasernas syntax. Söksträngen för databasen Scopus såg ut på följande sätt:²

TITLE-ABS-KEY (("covert channel" OR "steganograph*" OR "side channel*" OR "data tunnel*" OR "timing channel*" OR "data breach*" OR "exfiltrat*" OR "data leak*" OR "data theft*" OR "data stealing") AND ("detect*" OR "countermeasure*"))*

Antalet träffar per databas och totalt anges i Tabell 3. Därtill anges hur många av träffarna som kvarstod när dubletter inom varje databas strukits.

² Söktermen "steganograph*" ingick i söksträngen men steganografi exkluderades sedan ämnesmässigt eftersom det insågs att de allra flesta artiklar om det låg utanför fokus för denna litteraturgenomgång (ungefär som ren kryptering också gör det). Däremot inkluderas artiklar om vissa nätverksbaserade dolda kanaler vilka kan ses som en del av steganografi.

Tabell 3: Antal träffar i de olika databaserna samt hur många träffar som var unika.

Databas	Träffar	Träffar exkl. dubletter
Scopus	7613	7537
IEEE Xplore	2547	2535
ACM Guide to Computing Literature	488	483
Science Direct	431	398
<i>Totalt</i>		<i>10953</i>

Utöver sökning i dessa databaser identifierades artiklar också genom sökning bland referenser och citeringar, så kallad snöbollning (eng. snowballing) i enlighet med Kitchenham och Brereton (2013). På så vis inkluderades också relevanta referenser (snöbollning bakåt) till de artiklar som identifierats i databassökningen och som klarat relevanskriterierna i avsnitt 4.2. Därtill inkluderades även relevanta artiklar som citerade artiklarna (enligt Google Scholar), i snöbollning framåt. Snöbollning kan upprepas för de nya artiklar som identifierats utanför databaserna. I denna genomgång bedömdes dock att det inte tillkom tillräckligt många artiklar för att göra mer än en första omgång snöbollning nödvändig.

4.2 Kontroll av artiklarnas relevans

En kombination av halvautomatisk och manuell exkludering gjordes enligt Tabell 4. Exkluderingskriterierna togs fram som en kombination av några av de som rekommenderades av Kitchenham och Charters (2007) samt några av de som användes i genomgångarna i avsnitt 3.2. Exempelvis hade genomgången Sabir m.fl. (2022) ett kriterium snarlikt det om lågt citeringstal som används här.

Tabell 4: Exkludering av artiklar.

Exkludering	Strukna	Kvar
Ingen	0	10953
Dubletter mellan databaser (enligt titel)	2884	8069
Konferensbeskrivningar (conference review samt keynotes)	459	7581
Editorial, erratum, note, letter, short survey, ebook, preprint	20	7561
Book, book chapter	127	7434
Trasig formatering	1	7433
Publicerade före 2019 med citeringstal under 5 (enligt metadata från Scopus respektive IEEE samt i övrigt enligt Scholar)	1955	5478
Irrelevant ämne utifrån titel (se nedan om ämnen)	5270	208

Exkludering	Strukna	Kvar
Irrelevant ämne utifrån sammanfattning	135	73
Irrelevant ämne utifrån fulltext	26	47
Överlapp med annan artikel av samma författare	3	44
Tillägg från snöbollning bakåt (referenser)	-7	51
Tillägg från snöbollning framåt (citeringar)	-4	55

Bedömning av titlarnas ämnesrelevans gjordes uppdelat på två personer (författarna). Eftersom antalet artiklar var så stort exkluderas, hellre än inkluderades, artiklar med tveksam relevans. Hundra av artiklarna bedömdes av båda författarna och detta visade på tillräcklig likhet för att fortsätta processen. Därefter gjordes en dubbelkontroll genom att författarna gjorde bedömningar (för eventuellt exkluderande) av varandras inkluderande artiklar och skillnader diskuterades. Därpå tog rapportens huvudförfattare beslut om slutlig inkludering eller exkludering.

För artiklar som inte redan exkluderats gjordes liknande bedömningar av sammanfattning och i nödvändiga fall även av fulltext. En skillnad gentemot bedömningarna av titel var att tveksamma artiklar hellre inkluderades än exkluderades. Artiklarna bedömdes som ämnesmässigt relevant eller irrelevant enligt Tabell 5, utgående från forskningsfrågorna och deras avgränsning.

Tabell 5: Ämnesmässig inkludering och exkludering.

Inkluderat (ämnesmässigt)	Exempel på exkluderat
Nätverksbaserad exfiltrering från it-system	Hårdvarubaserade sidokanaler samt fysiska angrepp.
Detektion när exfiltreringen sker	Exfiltreringstekniker snarare än deras detektion samt försvar som baseras på prevention (t.ex. åtkomstkontroll och informationssäkerhetsklassning) eller utredning efteråt (t.ex. forensik).
Antagonistisk exfiltrering	Metoder för att hindra misstag där personer råkar skicka mejl till fel person.
Exfiltrering snarare än exempelvis styrning av skadlig kod	Tekniker för att styra skadlig kod (Command and Control), tekniker för att som användare hämta information in till nätet samt steganografi (vilket ofta fokuserar på döljande i vila etc. snarare än vid nätverksöverföring).

Efter alla exkluderingskriterier applicerats kvarstod alltså 55 artiklar. Det kan noteras att av de slutligt inkluderade artiklarna fanns alla utom två (Bubnov 2018 samt Salih m.fl., 2015) med i Scopus och dessa två saknades även i de andra databaserna (men identifierades vid snöbollningen i Scholar).

De 55 artiklarna delades upp i två typer: A) 48 artiklar som bygger upp en experimentmiljö och experimentellt utvärderar en viss detektionsmetod; samt B) 7 genomgångar av sådana artiklar. Typ A utgör de artiklar som är stommen för denna litteraturgenomgång och dess resultat. Typ B inkluderas för att med snöbollning hitta artiklarna av Typ A samt för att bidra till identifieringen (i avsnitt 3.2) av vad det finns för tidigare arbeten som liknar denna rapports genomgång.

4.3 Extrahering av data

Artiklarna (med experimentdata) delades upp på två personer och varje artikel bedömdes av en person, med en dubbelkontroll för ungefär en tredjedel av artiklarna. För varje artikel tog det cirka 15 minuter att extrahera data enligt Tabell 6. Kategoriseringen och skalorna togs fram av rapportens ena författare utgående från tidigare genomgångars kategoriseringar tillsammans med anpassningar efter pilotextraheringar.

Tabell 6: Kategorisering för extrahering av data.

Kategori	Data	Skala
Bibliografi	Publikationstyp	Tidskrift eller Konferens
	Publikationsår	2012–2022
	Citeringar	0 och uppåt
Kvalitet på ansatsen	Tydligt syfte	Ja/Nej samt fritext
	Tydliga forskningsfrågor	Ja/Nej samt fritext
	Tydlig och lämplig metod	Ja/Nej samt fritext
Miljö	Legitima data	Live, Insamlade, Syntetiska samt Egna eller från Referens
	Nätverkets typ (organisation, teknisk uppbyggnad och om skarpt nätverk eller labbmiljö)	Fritext
	Nätverkets storlek (antal användare, datamängd och tid det studerades)	Fritext

Kategori	Data	Skala
Angreppsscenario	Data som exfiltreras	Fritext
	Tänkt angripare	Fritext
Teknik för exfiltrering	Döljande	Exempelvis tunnling, kryptering, vilseledning
	Nätverksprotokoll	Exempelvis HTTP, HTTPS, DNS, IP
	Illegitim trafik (t.ex. exfiltreringsverktyg eller dataset från referens)	Fritext
Metod för detektion	Kännetecken som detektionen baseras på (t.ex. mängden trafik)	Fritext
	Typ av detektionsalgoritm	Signaturbaserad, anomalibaserad, traditionell maskininlärningsbaserad eller djupinlärningsbaserad
	Detektionsalgoritm	Fritext
	Utvärderingsmått (t.ex. sannolikheten för falska positiva eller tid tills detektion)	Fritext
	Kvaliteten på utvärderingen av metoden samt metodens nackdelar	Fritext
Övriga kommentarer	Övrigt	Fritext

Extraherade data skrevs ner per artikel i ett extraheringsark i ett kalkylprogram. I extraheringsarket fanns förutom fält enligt tabellen också ett fält för övriga kommentarer där de som utförde extraheringen kunde lägga till förtydliganden och andra kommentarer till hjälp för litteraturgenomgången.

4.4 Syntes av data

För att skapa en syntes av data användes dels statistik (för kvantitativa data), dels kategorisering och tematisk analys (för kvalitativa data). Den statistiska analysen räknade exempelvis antal artiklar av olika kategorier utan några komplicerade statistiska mått. Den tematiska analysen hittade lämpliga kategoriseringar av det kvalitativa materialet. Kategoriseringarna vägledades av tidigare forskning och framförallt då tidigare litteraturgenomgångar inom området. I vissa fall fanns redan lämpliga färdiga kategoriseringar i sådana källor vilka då användes. För att säkerställa att alla kategoriseringar var lämpliga gjordes flera mindre pilotförsök med iterationer av extraheringar och synteser tills extraheringskategorierna sågs som lämpliga och synteserna gav lämpliga sammanställningar. Till slut gjordes själva sammanställningarna.

5 Artiklarnas ursprung och kvalitet

I detta avsnitt beskrivs artiklarnas ursprung i form av var och när de publicerades. Därtill beskrivs artiklarnas kvalitet objektivt utifrån hur många citeringar de fått av andra artiklar samt subjektivt utifrån en bedömning av artiklarnas syften och metoder.

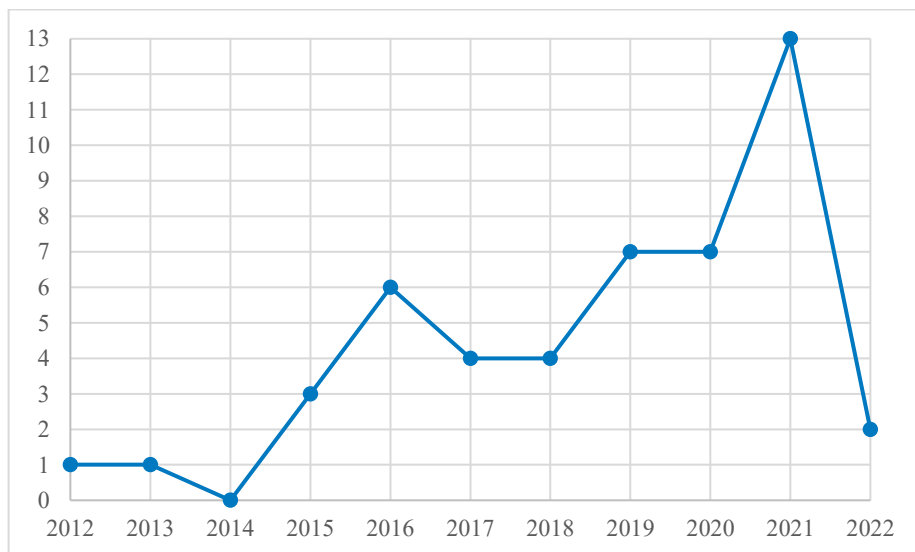
5.1 Ursprung

Tabell 7 anger huruvida artiklarna publicerats i en tidskrift eller i konferenssammandrag.

Tabell 7: Antal och andel artiklar som publicerats i tidskrift eller i konferenssammandrag.

Typ av publikation	Antal	Andel
Tidskrift	18	38 %
Konferens	30	62 %

Figur 3 anger när artiklarna publicerades. Notera att sökningen gjordes tidigt under 2022 varför antalet intressanta artiklar som publiceras under året förmodligen kommer öka.



Figur 3: Antal artiklar som publicerats per år 2012–2022.

5.2 Citeringar

Artiklarna har i medeltal citerats 13 gånger av andra artiklar. Citeringarna är dock väldigt ojämnt fördelade med två artiklar som når väldigt höga siffror (98 respektive 166 citeringar); 18 artiklar med mellan 11 och 55 citeringar; samt 28 artiklar med mellan 0 och 10 citeringar. En artikels antal citeringar kan öka med tiden och det är därför naturligt att yngre artiklar inte hunnit få så många citeringar, vilket framgår av Tabell 8. Därtill bör det noteras att ett exkluderingskriterium rörde antal citeringar och då bara för artiklar publicerade före 2019. Utan kriteriet hade äldre artiklar med lägre citeringsgrad kunnat komma med i urvalet och därmed kunna dra ner medeltalet för tidigare år.

Tabell 8: Antalet citeringar per artikel och år.

År	Citeringar per artikel i genomsnitt
2012	24
2013	55
2014	–
2015	62
2016	32
2017	21
2018	17
2019	12
2020	9
2021	2
2022	2

5.3 Subjektiv kvalitetsbedömning

Varje artikels syfte är tydligt i termer av vilket ämnet är och varför forskningen görs. I de artiklar som ingick i genomgången var syftet oftast att ta fram en ny detektionsmetod och att utvärdera den. När det görs är det ofta otydligt vad det specifika målet är och experiment görs nästan alltid utan tydliga forskningsfrågor eller hypoteser. Enbart i två fall (Sherry m.fl., 2015; Shvartzshnaider m.fl. 2019) finns tydliga forskningsfrågor och då enbart i samband med den experimentella utvärderingen av detektionsmetoden snarare än redan i början av artikeln.

Forskningsfrågorna är bland annat på den öppna formen *"Hur få exfiltreringar missar vår detektionsmetod?"* eller *"Hur stor blir overheaden på grund av vår detektionsmetod?"*. I bara tre fall (Cabaj m.fl., 2018; Davis och Foo, 2016; Ishikura m.fl., 2021) finns en hypotes och då enbart som del av löptexten (t.ex.

"Vi har som hypotes att antal förfrågningar uppfyller Zipfs lag", fritt översatt) och utan att det är tydligt huruvida hypotesen är artikelns kärna. Hypotesen besvaras sedan (helt eller delvis bekräftande) i förbifarten.

Artiklarnas metodbeskrivningar är för det mesta begränsat beskrivna vilket gör det svårt eller omöjligt att extrahera vissa data från artiklarna och förstå forskningen. När beskrivningen är tillräcklig för att bedöma om metoden är vettig framkommer det att det finns stora variationer. Ingen artikel är dock helt dålig eller helt bra på alla punkter. I värsta fall beskrivs ett experiment med en eller flera begränsningar i form av mycket förenklade och små labbmiljöer, orealistiska data som ska exfiltreras, angriparscenarion som knappast är rimliga samt knapphändiga utvärderingar av för- och nackdelar. I bästa fall är experimentet upplagt med flera av följande egenskaper: stora miljöer fyllda av riktig legitim trafik i riktiga nätverk eller som spelats in, där exfiltrerade data och angreppsscenarion baseras på riktiga angrepp samt där utvärderingarna ser på problemet från olika håll och väger olika utfallsmått mot varandra. Enbart två artiklar (Ahmed m.fl., 2020; Bortolameotti m.fl., 2017) nämner att de etikprövat sin forskning. Artiklarna ger inga detaljer om varför etikprövningen gjordes men eftersom de använder data från nätverk med riktiga användare gjordes förmodligen prövningarna för att se till att studierna av användarna skedde på ett etiskt lämpligt sätt.

6 Miljön för exfiltreringen

Detta kapitel beskriver varifrån artiklarnas legitima data kommer, det vill säga den nätverkstrafik som skymmer hotaktörernas exfiltreringsdata. Kapitlet beskriver också vilken typ av nätverk som används i experimenten, hur stort nätet är samt hur mycket trafik det innehåller (före exfiltreringen).

6.1 Legitima datas ursprung

Tabell 9 beskriver varifrån artiklarna fått sina legitima data. Data kan delas in i tre typer: live, insamlade och syntetiska. *Live* betyder att data från riktiga nätverk används direkt medan *Insamlade* betyder att riktiga nätverk visserligen används men att data spelas in där för att sedan spelas upp i experimentmiljön. *Syntetiska* betyder att data tillverkas i labbmiljö, oavsett om det sker på realistiskt sätt som i ett vanligt nätverk eller helt konstruerat. Därtill finns det artiklar som använder data från en referens (det vill säga en annan artikel). Den här litteraturgenomgången har i sådana fall extraherat huvudartikelns beskrivning av datasetet utan att göra några försök att hitta dess referens och extrahera mer information om datasetet därifrån. Notera att data som är live aldrig kan komma från en referens. Tre artiklar använder data från flera ursprung.

Tabell 9: Hur många artiklar som får sina legitima data från varje källa.

Legitima datas ursprung	Antal
Syntetiska (exkl. de från referens)	14
Syntetiska (via referens)	2
<i>Syntetiska (totalt)</i>	<i>16</i>
Insamlade (exkl. de från referens)	16
Insamlade (via referens)	5
<i>Insamlade (totalt)</i>	<i>21</i>
<i>Live</i>	<i>4</i>
<i>Ej specificerade (men från referens)</i>	<i>7</i>
<i>Ej specificerade (exkl. de från referens)</i>	<i>3</i>

För insamlade legitima data från riktiga nätverk kan det komma med illegitima data, det vill säga spår av naturligt förekommande angrepp. I ett mindre antal artiklar nämns försök att tvätta bort sådana spår för att enbart ha kvar legitima data. Tvättningen består av att ta bort trafik som rör domäner som är ovanliga på internet (som i Yang m.fl., 2020); av att applicera detektionsmetoden i förväg (som i Preston, 2019); av att göra statistiska jämförelser med ett annat dataset

(som i Steadman och Scott-Hayward, 2018); eller av att kontrollera URL:er mot listan på skadliga sådana på Virustotal (som i Zhan m.fl., 2022).

6.2 Nätverken

Tabell 10 beskriver vilken typ av nätverk som används i experimenten. Hur nätverken beskrivs i artiklarna varierar, vilket återspeglas här. Vissa nätverk beskrivs utifrån vem som äger det eller vilken organisations nätverk det ska likna. Andra nätverk beskrivs utifrån hur det är uppbyggt tekniskt eller huruvida det används i skarpt läge eller labbmiljö. Tre artiklar använder flera typer av nätverk. Därtill är det ganska vanligt att artiklarna inte beskriver sina nätverk, därav den ganska höga siffran för *ej specificerat*.

Tabell 10: Typen av nätverk som artiklarna använder i sina experiment. Antalet artiklar framgår av kolumnen #.

Typ av nätverk	#	Artiklar (exempel)
Labbmiljö	17	Chiu m.fl., 2019; Chowdhary m.fl., 2021
Internetleverantörsnätverk	2	Nadler m.fl., 2019; Yang m.fl., 2020
Universitetsnätverk	9	Ahmed m.fl., 2020; Al-Bataineh och White, 2012
Forskningsinstitutsnätverk	1	Ahmed m.fl., 2020
Internationell organisations nätverk	1	Bortolameotti m.fl., 2017
Kommunalt nätverk	1	Tsikerdekis m.fl., 2021
Företagsnätverk	6	D'Angelo m.fl., 2022; Durkota m.fl., 2017
P2P-nätverk (labbmiljö)	1	Bubnov, 2018
Mobilnät med 2G, 3G, 4G (labbmiljö)	1	Do m.fl., 2017
SDN (labbmiljö)	1	Cabaj m.fl., 2018
Taktiskt militärt nätverk (labbmiljö)	1	Aurisch m.fl., 2021
Ej specificerat	10	-

Tabell 11 beskriver nätverkens storlek i form av antal användare, storlek på datamängd och hur lång tid de studerades. Det finns stora variationer i samtliga tre aspekter. Förmodligen leder detta till stora svårigheter med att jämföra artiklarnas resultat, där vissa kan uppvisa gediget utvärderade detektionsmetoder medan andra enbart kan uppvisa mycket preliminära resultat. Därtill är det vanligt att inte beskriva de aspekter som nämns i tabellen, därav de höga siffrorna för *ej specificerat*.

Tabell 11: Artiklarnas data- och nätverksstorlekar samt varaktigheten i studierna av nätverken. Antalet artiklar framgår av kolumnen #.

Data- och nätverksstorlek samt varaktighet	#	Artiklar (exempel)
100 000 användare eller mer	1	Nadler m.fl., 2019
1000–10 000 användare	5	Aiello m.fl., 2016; Davis och Foo, 2016
100-1000 användare	1	Ellens m.fl., 2013
10-100 användare	2	Al-Bataineh och White, 2012; Ishikura m.fl., 2021
1-10 användare	5	Almusawi och Amintoosi, 2018; Salih m.fl., 2015
Ej specificerat	34	–
17 GB data	1	Dijk, 2021
1,5 GB data	1	D'Angelo m.fl., 2022
100 miljoner+ paket	1	Chen m.fl., 2021
10–50 miljoner paket	2	Iglesias m.fl., 2016; Lambion m.fl., 2020
100 000–500 000 paket	2	Altuncu m.fl., 2021; Wendzel m.fl., 2019
1–10 000 paket	1	Steadman och Scott-Hayward, 2018
500 miljarder förfrågningar	1	Yu m.fl., 2016
5 miljarder förfrågningar	1	Nadler m.fl., 2019
30 miljoner förfrågningar	1	Preston, 2019
1–20 000 förfrågningar	2	Cabaj m.fl., 2018; Chowdhary m.fl., 2021
1 miljon flöden	1	Vekshin m.fl., 2020
150 000 flöden	1	Ding m.fl., 2021
51 580 mejl	1	Cui m.fl., 2021
Ej specificerat	32	–
1–3 månader	3	D'Angelo m.fl., 2022 Durkota m.fl., 2017;
1 vecka–1 månad	5	Ishikura m.fl., 2021; Nadler m.fl., 2019
24–48 h	2	Al-Bataineh och White, 2012 Ellens m.fl., 2013;
Ej specificerat	38	–

7 Angreppsscenario för exfiltreringen

Detta kapitel beskriver vilka typer av data vars exfiltrering artiklarna studerar. Därtill beskrivs vilka typer av hotaktörer som artiklarna antar utför exfiltreringen.

7.1 Data som exfiltreras

Tabell 12 beskriver vilka data som exfiltreras i artiklarna. I de flesta artiklar saknas sådan information och när information finns är den knapphändig.

Tabell 12: Data som exfiltreras i artiklarnas experiment. Antalet artiklar framgår av kolumnen #.

Data som exfiltreras	#	Artiklar
1–2 MB stora filer	1	Ellens m.fl., 2013
Inloggningsuppgifter för distansmöten	1	Shvartzshnaider m.fl., 2019
Kreditkortsdetaljer	2	Ahmed m.fl., 2020; Das m.fl., 2017
Personlig info inklusive kreditkortsdetaljer	1	King m.fl., 2021
Slumpdata (för att likna krypterade data)	1	Wendzel m.fl., 2019
Känsliga dokument tagna från Wikileaks	1	Bortolameotti m.fl., 2017
Ej specificerat	41	-

7.2 Hotaktörerna

Tabell 13 beskriver vilka typer av hotaktörer som artiklarnas experiment anpassas efter. Hotaktörerna beskrivs sällan alls i artiklarna och när så görs är det knapphändigt. Ett vanligt sätt att beskriva hotaktörer är utifrån deras intention, tillfälle och kapacitet. Utifrån detta kan det om beskrivningar av hotaktörerna sägas följande:

- För *intentionen* är informationen särskilt begränsad vilket gör det svårt att avgöra hur sannolikt angrepp är och vad konsekvenserna kan bli. Å andra sidan är det rimligt att varje försvarare gör sin egen avvägning mellan risken och de tolererbara riskreduceringskostnaderna.
- För *tillfället* finns det ibland information om huruvida hotaktören är en insider men det saknas information om när systemet är närbart.
- För *kapaciteten* beskrivs den ibland uttryckligen i termer av en avancerad och ihärdig hotaktör (eng. Advanced Persistent Threat, APT). Kapaciteten beskrivs även implicit genom att hotaktörens kapacitet kan

förstås utifrån vilken exfiltreringsteknik som används. Sådan implicit information tas inte upp i detta avsnitt.

Tabell 13: Vilka typer av hotaktörer som artiklarnas experiment anpassas efter. Antalet artiklar framgår av kolumnen #.

Hotaktör	#	Artikel
Både insiders och outsiders. Angriparen förnimmar försvararens strategi och vet ibland strategin exakt. Angriparen kan normalt exfiltrera genom att lägga till trafik men i vissa situationer även ersätta legitima trafikdata med exfiltreringsdata och därmed exfiltrera mer utan att det ser konstigt mycket ut.	1	Durkota m.fl., 2017
En angripare som (förstås) inte har kontroll över detektionsmetodens användning.	1	Bortolameotti m.fl., 2017
En insider.	1	Chung m.fl., 2020
Ett botnät baserat på Zeus vilket är benäget att stjäla data.	1	Al-Bataineh och White, 2012
APT:n Wekby.	1	Das m.fl., 2017
En (tänkt) APT som är svårдетекterad eftersom den är avancerad och lågintensiv.	1	Marchetti m.fl., 2016
Ej specificerat	42	

Artiklarnas information om hotaktörerna beskrivs typiskt före experimentresultaten. I ett fåtal fall (Bortolameotti m.fl., 2017; Cui m.fl., 2021; Das m.fl., 2017; Ishikura m.fl., 2021) beskrivs hotaktören (även) efteråt och då i termer av hur angriparen kan tänkas anpassa sig för att gå runt detektionsmetoden. Sådan information i efterhand ger inte insikter i typen av hotaktör men indikerar ett självkritiskt förhållningssätt vilket är viktigt för att täppa till brister i skydden och stoppa hotaktörer i praktiken.

8 Dataexfiltreringsteknikerna som ska detekteras

Detta kapitel beskriver hur dataexfiltreringen i artiklarna går till och därmed vad detektionsmetoderna (som beskrivs i nästa kapitel) ska detektera. Närmare bestämt beskrivs hur exfiltreringen döljs och vilka protokoll som används för exfiltreringen samt vilka verktyg och dataset som används.

8.1 Exfiltreringens döljande och protokoll

Tabell 14 nämner vilka sätt som används i artiklarna för att dölja exfiltreringen. Detta är alltså hur (den tänkta) hotaktören försöker undgå att försvararen detekterar att exfiltrering sker eller åtminstone vad som exfiltreras. Därtill beskriver tabellen vilka protokoll artiklarna använder för att exfiltrera.

En artikel kan använda flera av sätten att dölja samtidigt varpå en artikel räknas med i flera rader i tabellen. Det kan dock finnas sätt att dölja som används i artiklarna men som inte syns här. Det gäller framförallt döljande som ingår i funktioner i de exfiltreringsverktyg som artiklarna använder (se avsnitt 8.2). Därtill justerar artiklarna ibland exfiltreringens intensitet för att hamna under vad som skulle bli detekterat. Exempelvis valde Cabaj m.fl. (2018) att bara exfiltrera en bit per paket för att försvåra för detektionsmetoden trots att exfiltreringstekniken möjliggjorde att åtta bitar per paket exfiltrerades. Sådant döljande via låg intensitet syns inte i tabellen eftersom det rör sig mer om en kalibrering av en teknik än ett helt eget sätt att dölja.

Tabell 14: Artiklarnas sätt att exfiltrera i form av döljande och protokoll. Notera att kombinationer av döljande inte syns i tabellen. Antalet artiklar framgår av kolumnen #.

Döljande	Protokoll	#	Artiklar (exempel)
Tunnling	DNS över HTTPS	5	Mitsubishi m.fl., 2021; Zhan m.fl., 2022
	DNS över övriga protokoll såsom FTP och HTTP	1	Almusawi och Amintoosi, 2018
	SSH, P2P över DNS	1	Aiello m.fl., 2016
	SCP, SSH, IMAP, POP3, SMTP över HTTP	1	Davis och Foo, 2016
	SSH, RDP över DNS, HTTP, HTTPS	1	Lin m.fl., 2019

Döljande	Protokoll	#	Artiklar (exempel)
Data på ovanligt ställe (i pakethuvud och/eller i protokoll som normalt inte används för överföring av användardata)	DNS	22	Buczak m.fl., 2016; Lambion m.fl., 2020
	TCP/IPv4	7	Iglesias m.fl., 2016; Marchetti m.fl., 2016
	TCP/IPv6	1	Salih m.fl., 2015
	HTTP, SMTP, IMAP	7	King m.fl., 2021
Data döljs som tid mellan paket (tidskanal)	TCP	1	Al-Eidi m.fl., 2021
Kryptering (filkrypto)	HTTP	3	Al-Bataineh och White, 2012; Bortolameotti m.fl., 2017
	DNS	1	Ahmed m.fl., 2020
	Ej specificerat	1	Wendzel m.fl., 2019
Kryptering (transportkrypto i protokollet)	HTTPS	8	Sherry m.fl., 2015
Angriparen ersätter legitim trafik och undgår därmed volymgräns	Google drive upload	1	Durkota m.fl., 2017
Ej specificerat	Mejl	2	Chung m.fl., 2020; Shvartzshnaider m.fl., 2019
	HTTP (webbmejl)	1	Cui m.fl., 2021
	FTP och google drive upload	1	Dijk, 2021

Det bör noteras att de protokoll som nämns här främst rör de protokoll där data som ska exfiltreras läggs in. Det protokollets paket kommer dock inkapslas i andra protokoll i flera lager. Exempelvis är DNS ett protokoll på hög nivå och DNS-paketen skickas därför typiskt inuti UDP-paket vilka i sin tur typiskt skickas inuti IP-paket och så vidare ner till och med det fysiska lagret. Även artiklarna är knapphändiga med information om protokollen på undre lager.

Med tunnling döljs paket genom att de kapslas in i andra paket på ett sätt som avviker från den normala inkapslingen i nätverket. Ett exempel på tunnling är att exfiltrera data i DNS-paket vilka skickas inbakade i HTTPS-paket.

Kombinationen DNS över HTTPS (känd som DoH) är sedan några år tillbaka standardiserad och används av legitim DNS-trafik men kan också utnyttjas av

illasinnade. Om angriparen vill exfiltrera via DNS men försvararen har en intern DNS-server (eller brandvägg) som kan detektera exfiltrering via DNS måste angriparen göra två saker för att möjliggöra sin exfiltrering. Först måste angriparen hitta en extern DNS-server som inte detekterar exfiltrering via DNS.³ Sådana finns det gott om eftersom internet bygger på att DNS-paket behandlas utan censurering. Sedan måste angriparen använda DoH så att DNS-paketen kapslas in i HTTPS och inte läses av den interna DNS-servern. På så vis tunnlas DoH-paketen ut ur det interna nätet till en punkt där vanlig DNS-exfiltrering fungerar.⁴

8.2 Verktyg och dataset

Tabell 15 beskriver vilka verktyg och dataset som artiklarna använder för att exfiltrera data. Dataset utgör färdiga data med exfiltreringstrafik och som typiskt är allmänt tillgängliga. Det finns tre kategorier av verktyg förutom dataset. Det överlägset vanligaste alternativet är att använda färdiga verktyg som andra forskare tagit fram för att forska om detektion av exfiltrering eller om andra säkerhetsfunktioner. De vanligaste av dessa verktyg är Iodine, Dnscat2, Dns2tcp och DET. Därtill är det ganska vanligt att en artikel använder flera av dessa. Det näst vanligaste alternativet är egenframtagna verktyg där en artikels forskare själva tagit fram ett verktyg. Därtill har några artiklar använt riktig skadlig kod som blivit allmänt tillgänglig. Dessa verktyg körs med försiktighet i labbmiljö i studierna. Det fjärde alternativet är som sagt att använda färdiga data från en annan källa. En sammanställning av angrepp som ingår i olika dataset (vilka inte nödvändigtvis ingår i den här litteraturgenomgångens artiklar) finns i tabell iv i Ring m.fl. (2019). När färdiga dataset används kan dessa innehålla både exfiltreringsdata och legitima data. På motsvarande sätt som i avsnitt kan legitima data då tvättas bort. Sådan tvättning nämns i en artikel (Lambion m.fl., 2020) och utgörs där av att låta experter bedöma om data är legitima.

³ För att skicka paket till den externa DNS-servern kan det krävas att den interna DNS-servern hjälper till med ett första vanligt DNS-uppslag. Därför behöver den externa DNS-servern typiskt vara legitim i den interna DNS-servrens ögon.

⁴ Se även Jacobs (2018). Därtill påminner tillvägagångssättet i DoH om domänfrontning där HTTPS-paket har en yttre öppen adress som leder paketet till en server samt en inre krypterad adress som bara servern kan se och får den att (okänt för utomstående) skicka vidare paketet till en annan mottagare i samma nätverk (se Mitre, 2020a).

Tabell 15: Verktyg och dataset (från andra källor) som artiklarna använder för att exfiltrera data. För färdiga verktyg nämns enbart de vanligaste (tillsammans med när de senast uppdaterades). Antalet artiklar framgår av kolumnen #.

Typ	Verktyg	#	Uppd	Beskrivning
Färdiga verktyg	Iodine ⁵	14	2022	Tunnlar data i IPv4 över DNS. Har ej kryptering.
	Dnscat2 ⁶	12	2022	Skickar data över DNS. Har kryptering.
	Dns2tcp ⁷	12	2017	Tunnlar data i TCP över DNS. Har ej kryptering.
	DET ⁸	4	2019	Skickar data över många protokoll såsom HTTP, HTTPS, ICMP, DNS, SMTP, IAMP, TCP, UDP, FTP. Har kryptering.
	Övriga färdiga verktyg	30	-	-
Övriga verktyg samt dataset	Egenframtaget verktyg	11	-	Exempelvis lägga in ett extra ord av fix längd i varje DNS-förfrågan för att inte någon cache ska kunna besvara förfrågan på vägen till hotaktörers DNS-server (se Preston, 2019)
	Riktig skadlig kod	4	-	Datastjälande botnätet Zeus (i Al-Bataineh och White, 2012); åtta familjer skadlig kod (i Bortolameotti m.fl., 2017); skadlig kod som nyligen använts av en APT (i Das m.fl., 2017); samt FrameworkPOS och Backdoor.Win32.Denis (i Nadler m.fl., 2019)
	Dataset från annan källa	6	-	(Se t.ex. Ding m.fl., 2021; Dijk, 2021)
	Ej specificerat	13	-	-

⁵ Iodine har öppen källkod, se Github (2022a).

⁶ Dnscat2 har öppen källkod, se Github (2022b).

⁷ Dns2tcp har öppen källkod, se Github (2017).

⁸ DET har öppen källkod, se Github (2019).

9 Metoder för att detektera illasinnad dataexfiltrering

I detta kapitel beskrivs olika metoder för att detektera illasinnad dataexfiltrering. Det beskrivs vad detektionsmetoden har för indata (kallade kännetecken) samt vilka algoritmer som metoden använder för att bedöma indata. Därtill beskrivs artiklarnas utvärderingar av metoderna.

Först följer här en kort beskrivning av hur artiklarna samlar in indata till algoritmen. Det typiska är att det finns en central nod i nätet där avlyssning sker med något typiskt loggverktyg för nätverk såsom Wireshark och Zeek. I ett fall (Aurisch m.fl., 2021) är nätet mer distribuerat varför varje enskild nod har en agent på sig som står för insamlingen.

9.1 Kännetecken som detektionen baseras på

Tabell 16 kategoriserar de kännetecken (eng. features) som detektionen baseras på i artiklarna. Kännetecknen utgör de variabler som detektionen använder som indata. Notera att varje kategori av kännetecken kan handla om både de data som exfiltreras (innehållet) och olika aspekter ur miljön runt exfiltreringen (kontexten). Ofta delar forskningsartiklarna upp kännetecken i just innehåll och kontext. Samtidigt finns det ingen entydig definition av begreppen varvid uppdelningen i innehåll respektive kontext beror på vilket synsätt som anammats. Det går till exempel att se allt som ligger i ett protokolls datafält som innehåll. Å andra sidan kan data exfiltreras inte bara i datafält utan även i protokolls pakethuvudfält. Det går också att se hela det översta lagrets paket (t.ex. DNS-paketet) som innehåll medan lägre lager blir kontexten. Därtill kan övrig inverkan på nätverket ses som kontext, t.ex. de effekter ett stort paket får. Innehållsbaserat kännetecken kan också enbart röra de data som en människa är intresserad av och som analyseras på ett sätt som liknar det sätt en människa tolkar data på. Endast en artikel (Shvartzshnaider m.fl., 2019) använder denna sista typ av tolkning av begreppet innehåll.

Det bör också noteras att de olika kategorierna är på olika nivå och tillsammans ger ett antal olika skärningar. Till exempel är teckenfrekvenser av domännamnet en del av kategorin entropi samtidigt som domännamnet är en del av kategorin datainnehåll. Kategoriseringen här har inspirerats av de gjorda i Sabir m.fl. (2022) samt Wendzel m.fl. (2015).

Tabell 16: Kännetecken som detektionen baseras på. Antalet per kategori står i parentes i kategori-kolumnen. Antalet artiklar framgår av kolumnen #.

Kategori (#)	Kännetecken	#	Artiklar (exempel)
Entropi (20)	Teckenfrekvenser i olika paketfält (såsom DNS-domännamnet)	11	Yang m.fl., 2020
	Entropi i övrigt (för större delar av paketet eller hela paketet)	9	Almusawi och Amintoosi, 2018
Tid (16)	Tid mellan paket	9	Aurisch m.fl., 2021
	Flödets varaktighet	2	Ellens m.fl., 2013
	Tid mellan förfrågan och svar	1	Montazerishatoori m.fl., 2020
	Tid i övrigt	4	Chung m.fl., 2020
Stränglängd (15)	Längden av strängar i olika fält (såsom för domännamnet)	15	Altuncu m.fl., 2021
Trafikflöden (15)	Antal förfrågningar per domän	4	Chowdhary m.fl., 2021
	Kvot mellan A-förfrågningar och AAAA-förfrågningar i DNS	1	Aurisch m.fl., 2021
	Andel paket skickade respektive mottagna	2	Yang m.fl., 2020
	Trafikflöden i övrigt	8	Do m.fl., 2017
Storlek på paketet (13)	IP-pakets storlek	13	Almusawi och Amintoosi, 2018
Trafikvolym (9)	Antal paket	1	Chen m.fl., 2021
	Antal anslutningar	1	Marchetti m.fl., 2016
	Toppar i trafiken	1	Ellens m.fl., 2013
	Trafikvolym i övrigt	6	Steadman och Scott-Hayward, 2018
Pakethuvud-innehåll (9)	TTL-värdet i IP-huvudet	4	Cabaj m.fl., 2018
	Värdet av IP-options	1	King m.fl., 2021
	Sekvensnumrets utseende	1	Nafea m.fl., 2020
	Pakethuvudinnehåll i övrigt	3	Ayub m.fl., 2019
Datainnehåll (7)	Existensen av bilagor	1	Chung m.fl., 2020
	Information som inte förväntas komma tillsammans	1	Shvartzshnaider m.fl., 2019
	Typen av innehåll	3	Bubnov, 2018
	Domännamnet	1	Chen m.fl., 2021
	Förhållandet mellan olika paketdatafält	1	Lin m.fl., 2019

Kategori (#)	Kännetecken	#	Artiklar (exempel)
Ej specificerade (2)	-	2	-

Artiklarna har typiskt valt kännetecken grundat i en rimligt angiven motivering och förmodligen grundat i artikelförfattarnas expertis. Samtidigt kan det ifrågasättas om valda kännetecken verkligen är de bästa. Hur kännetecken ska väljas på bästa sätt har studerats i en artikel: Davis och Foo (2016). Där testas en metod att automatiskt välja kännetecken genom användning av en maskininlärningsalgorithm på träningsdata. Det noteras också att valet av kännetecken inte får bli alltför anpassat till specifika träningsdata. Annars kommer algoritmen bara fungera bra om den sedan används med data som är väldigt lika träningsdata. På liknande sätt behöver frågan som algoritmen ska svara på formuleras på rätt sätt. Det måste alltså noggrant definieras vad som räknas som illasinnad exfiltrering av data och vad som räknas som legitima data. Det kan noteras att det också finns algoritmer (vid namn neurala nätverk) som fungerar bra på rådata snarare än att det behöver bestämmas och utvinnas särskilda kännetecken. Avsnitt 9.2 beskriver bland annat sådana algoritmer.

9.2 Algoritmer för detektionen

Förmågan hos en detektionsmetod att detektera exfiltrering beror inte endast på vilka indata (kännetecken) den har tillgång till att analysera. Förmågan beror också på metodens beslutskriterier och därmed vilken algoritm som används. Tabell 17 beskriver de algoritmer som artiklarnas detektionsmetoder använder. Algoritmerna delas in i ett antal kategorier utifrån deras sätt att ta beslut.

I enklaste fall finns enkla regler som avgör om ett visst värde hos indata ska klassas som exfiltrering eller ej. Exempelvis kan en detektionsmetod ha en förutbestämd regel som säger att e-postmeddelanden med bilagor över tio megabyte utgör exfiltrering. För sådana metoder kan indata inte bara ses som kännetecken på exfiltrering utan till och med mer exakta signaturer för exfiltrering. Detta regel- och signaturtänk präglade tidigare forskningen (2006–2016) inom delområdet detektion av illasinnade DNS-tunnlar varefter ett mer komplext tankesätt tagit över (Wang m.fl., 2021). En rimlig slutsats är att tidigare forskning var av enklare slag och fokuserade på enklare typer av problem medan nyare forskning försöker lösa mer komplicerade problem där det inte är lika tydligt vad som är rätt och fel. Det är också rimligt att forskare använder avancerade metoder för att hitta nya signaturer vilka kan inkluderas i det praktiska försvaret.

Den nyare forskningen som undersöker mer komplexa beslutsmetoder undviker exakta förutbestämda signaturer till förmån för uppskattningar om godartat och

normalt beteende. Algoritmerna kan då vara anomalibaserade och detektera exfiltrering baserat på vad som ses som anomalier jämfört med det godartade och normala beteendet.

Den nyare forskningen kan också utgå från uppskattningar om både godartat och illasinnat beteende för att därefter göra en klassificering av indata. Dessa klassificeringsalgoritmer delas av Sabir m.fl. (2022) in i maskininlärningsalgoritmer av typen traditionell respektive typen djupinläring. Det kan noteras att även de anomalibaserade metoderna kan ses som maskininlärningsmetoder eftersom även de lär sig om indatas utseende.

Därtill finns algoritmer som kan ses som hybrider eftersom de använder en kombination av en annan algoritm och mänskliga beslut, baserade på exempelvis domänskunskap hos experter och antaganden om motståndarens agerande.

Tabell 17: Detektionsalgoritmer kategoriserade enligt Sabir m.fl. (2022) och rapportförfattarnas egna bedömningar. Antalet artiklar framgår av kolumnen #.

Kategori	Algoritm	#	Artiklar (exempel)
Djupinlärnings-baserade	Neurala nätverk och djupinläring	9	Montazerishatoori m.fl., 2020; Yu m.fl., 2016
Traditionell maskininlärnings-baserade	Beslutsträd (inkl. random forest)	12	Al-Bataineh och White, 2012; Lin m.fl., 2019
	SVM ⁹	9	Al-Eidi m.fl., 2021; Ayub m.fl., 2019
	Naive Bayes	6	Cui m.fl., 2021; Vekshin m.fl., 2020
	KNN	3	Chowdhary m.fl., 2021; Yang m.fl., 2020
Anomali-baserade	K-means	2	Almusawi och Amintoosi, 2018; Do m.fl., 2017
Signatur-baserade	Mönsterigenkänning (inkl. reguljära uttryck)	4	He m.fl., 2015; Sherry m.fl.m, 2015
Hybrid	Kombination människa-maskin	2	Chung m.fl., 2020; Marchetti m.fl., 2016
Övriga	-	11	-
Ej specificerat	-	7	-

De olika algoritmerna är komplicerade varför endast kortfattade och förenklade förklaringar ges här.

⁹ Sabir m.fl. (2022) ser vissa typer av SVM som anomalibaserade.

Neurala nätverk (NN) och djupinlärning är inspirerade av biologiska neurala nätverk som exempelvis människans hjärna. NN utgår från rådata utan att en människa valt kännetecken. Därefter processar NN rådata i flera omgångar, så kallade lager. De olika lagren motsvarar att neuroner i den mänskliga hjärnan har lite olika uppgifter och samarbetar med varandra. Antalet lager i NN är tillräckligt många för att lagren tillsammans kan ses som djupa och detta förklarar varför även termen djupinlärning används. Dessa algoritmer fungerar särskilt bra med mycket stora mängder indata.

Beslutsträd är algoritmer där det finns olika förgreningar och där det för varje förgrening ställs en fråga som avgör vilken gren som ska följas och därmed vilken den nästa frågan blir. Ett exempel finns i Figur 2 i början av metodkapitlet vad gäller beslut om vilka forskningsartiklar som ska ingå i denna rapports litteraturgenomgång. I komplexa situationer är det svårt att avgöra vilka frågor som ska ställas och hur många nivåer av frågor det ska finnas. Då kan flera beslutsträd läggas samman till en så kallad random forest, där beslutet tas ifrån deras gemensamma svar.

SVM, eller Supervised vector machines, är en algoritm som hittar den kategorisering som gör att de olika kategorierna ligger så långt som möjligt från varandra och därmed är så olika varandra som möjligt. I enklaste fallet finns bara två kännetecken varför problemet kan ses som att dra en linje mellan två grupper punkter i ett diagram med x- och y-axel.

Naive Bayes är en relativt enkel (naiv) algoritm. Algoritmen utgår från att alla kännetecken är oberoende av varandra, vilket är ett antagande som förenklar mycket men som ofta är tveksamt. Ett enkelt exempel är när det ska avgöras om en viss frukt är en banan vilken kan beskrivas från kännetecknen färg, diameter och längd. Algoritmen kommer värdera dessa kännetecken oberoende av varandra för att beskriva sannolikheten för att frukten faktiskt är en banan.

KNN, eller K nearest neighbour, är en algoritm som går ut på att varje ny datapunkt (av kännetecknen) får samma klass (illasinntat eller godartat) som innehas av den klass som majoriteten av de K närmaste punkterna (grannarna) har.

K-means är en algoritm som sorterar data i K kluster där varje kluster utgörs av de datapunkter som är mest lika varandra och samtidigt mest olika datapunkterna i de andra klustren.

Mönsterigenkänning utgår från förutbestämda regler eller mönster för att hitta kombinationer av tecken i textsträngar. Detta kan exempelvis användas för att hitta kreditkortsnummer, där man specificerar att algoritmen som används ska söka efter 16 siffror i följd. Reguljära uttryck är ett sätt att beskriva mönster som ska kännas igen.

Kombination människa–maskin utgår från att domänexperter har ett nära samarbete med en av algoritmerna. Exempelvis kan en algoritm göra en första grov bedömning av vad som kan vara exfiltrering varpå en människa kan göra en noggrannare kontroll av den resulterande mängden data.

Det bör noteras att vissa artiklar använder flera algoritmer. Algoritmerna kan då antingen 1) få ta hand om en egen avgränsad del av problemet eller 2) ta hand om samma del och jämföras för att hitta bästa algoritmen alternativt 3) ta hand om samma del och läggas samman (ensemble eller stacking) genom till exempel ett medelvärde.

9.3 Detektionsmetodernas effektivitet och nackdelar

När artiklarna utvärderar sina detektionsmetoder används i de flesta fall något eller flera av de mått som beskrivs i Tabell 18. Det bör dock noteras att artiklarna långt ifrån alltid definierar sina mått.

Beskrivningarna i tabellen utgår från att detektionsmetoden enbart försöker detektera exfiltrering medan frånvaron av exfiltrering inte ska ge någon detektion utan frånvaron av detektion. Detta kan jämföras med att ett brandlarm detekterar när det brinner men däremot inte uttryckligen detekterar frånvaron av brand.

Tabell 18: Mått som ofta används i utvärderingarna av detektionsmetoderna. Antalet artiklar framgår av kolumnen #.

Mått	Beskrivning	#
Sant Positiva (SP)	Antalet fall där detektionsmetoden detekterar exfiltrering när exfiltrering skett.	4
Sant Negativa (SN)	Antalet fall där detektionsmetoden inte detekterar exfiltrering när exfiltrering inte skett.	2
Falskt Positiva (FP)	Antalet fall där detektionsmetoden detekterar exfiltrering trots att exfiltrering inte skett.	9
Falskt Negativa (FN)	Antalet fall där detektionsmetoden inte detekterar exfiltrering trots att exfiltrering skett.	7
Sant Positiv Kvot (SPK)	$SPK = \frac{SP}{SP + FN}$ Sannolikheten att detektionsmetoden detekterar exfiltrering när exfiltrering skett. Kallas på engelska för True Positive Rate (TPR) eller Recall.	23

Mått	Beskrivning	#
Sant Negativ Kvot (SNK)	$SNK = \frac{SN}{SN + FP}$ Sannolikheten att detektionsmetoden undviker att detektera när exfiltrering inte skett. Kallas på engelska för True Negative Rate (TNR).	1
Falskt Positiv Kvot (FPK)	$FPK = \frac{FP}{SN + FP}$ Sannolikheten att detektionsmetoden felaktigt detekterar exfiltrering när exfiltrering inte skett. Kallas på engelska för False Positive Rate (FNR).	13
Falskt Negativ Kvot (FNK)	$FNK = \frac{FN}{FN + SP}$ Sannolikheten att detektionsmetoden missar att detektera trots att exfiltrering skett. Kallas på engelska för False Negative Rate (FNR).	3
Noggrannhet	$Noggrannhet = \frac{SP + SN}{SP + SN + FP + FN}$ Sannolikheten att detektionsmetoden detekterar exfiltrering när exfiltrering skett plus sannolikheten att detektionsmetoden inte detekterar exfiltrering när exfiltrering inte skett. Kallas på engelska för accuracy.	19
Precision	$Precision = \frac{SP}{SP + FP}$ Sannolikheten att detektionsmetoden har rätt när den visar på detektion.	18
F1	$F1 = 2 * \frac{Precision * SPK}{Precision + SPK} = \frac{2SP}{2SP + FP + FN}$ Ett harmoniskt medelvärde mellan SPK och Precision. Måttet väger alltså samman att exfiltrering detekteras (inga FN) och att annat än exfiltrering inte detekteras (inga FP).	16

Mått	Beskrivning	#
AUK	Arean under den kurva (AUK) som ges av förhållandet mellan SPK och FPK vid varje inställning av detektionsmetoden för att avväga mellan SPK och FPK. Kallas på engelska för Area Under Curve (AUC).	8
Ej specificerade	-	11

Artiklarna rapporterar för det mesta höga (och bra) siffror för mått som SPK, noggrannhet och precision samt låga (och bra) siffror för mått som FPK, FNK, FP, FN. I ett fåtal fall blir dock siffrorna sämre, som för Almusawi och Amintoosi (2018) (77 % SPK och 79 % precision) samt Bubnov (2018) (83 % noggrannhet). Med tanke på att forskning ofta försöker lösa komplicerade problem på nya sätt vore det inte förvånande om fler artiklars detektionsmetoder uppvisade lite sämre siffror. Det vore åtminstone logiskt om det var vanligare att en avvägning krävdes mellan att hålla nere falsklarmen (FP) och att undvika uteblivna riktiga larm (FN). Det är ungefär detta AUK utgår från (som dock också blir högt och bra i många artiklar).

Det finns också andra typer av mått som kan användas. Exempelvis föreslår Sabir m.fl. (2022) att de gängse utvärderingsmåten för maskininlärningsalgoritmer är undermåliga och att det vore bättre att använda mått som Matthew-korrelation (det vill säga Pearson-korrelation anpassad för binära variabler).

Detektionsmetodernas snabbhet är också ett relevant mått. Artiklarna är sällan tydliga med hur snabba detektionsmetoderna är men det finns undantag och här ges några exempel. I Altuncu m.fl. (2021) tog detektion av en DNS-tunnel under en millisekund. Marchetti m.fl. (2016) behövde två minuter för att analysera 140 miljoner flöden i en miljö på 10 000 maskiner. Ahmed m.fl. (2020) behövde för sin detektion en millisekund per DNS-förfrågan. Durkota m.fl. (2017) bedömde hur mycket data som hann exfiltreras innan detektion för en fix nivå falska positiva.

Generellt beskriver artiklarna inte några begränsningar med sina detektionsmetoder. Därtill är utvärderingen ibland tveksam (i Eremeev m.fl., 2021), tillrättalagd (i Ahmed m.fl., 2020), begränsad (i Ayub m.fl., 2019; Ellens m.fl., 2013; samt i Nafea m.fl., 2020) eller särskilt undermålig (Al-Bataineh och White, 2012; Bubnov, 2018; samt Do m.fl., 2017). Å andra sidan har de flesta artiklar utvärderat på en normal nivå. Det finns också de som utvärderat särskilt bra i form av Ding m.fl. (2021) samt Ishikura m.fl. (2021). Därtill är det rimligt om en artikel som fokuserar på att föreslå en ny detektionsmetod inte också hunnit utvärdera den perfekt.

10 Diskussion

Litteraturgenomgångens resultatdel har beskrivit vad forskningsartiklar fokuserat på inom området. Att diskutera dessa resultat utan särskilda referensramar bedöms vara av begränsad nytta. I detta diskussionskapitel sätts därför resultaten i relation till 1) hur litteraturgenomgången utförts och vilka resultat andra genomgångar fått; 2) vilka praktiska implikationer som finns vad gäller militära miljöer, riktiga angrepp, produkter för detektion, och alternativ till detektion; samt 3) vilken ytterligare forskning som skulle kunna utföras.

10.1 Litteraturgenomgångens kvaliteter samt skillnader mot andra genomgångar

Systematiska litteraturgenomgångar (SLR) är kraftfulla men mödosamma. De ger en mer objektiv bild av litteraturen än en icke-systematisk genomgång skulle göra. Å andra sidan har även SLR:er brister. Vilka dessa brister kan vara för denna rapports genomgång beskrivs närmare nedan. Därtill beskrivs hur denna genomgång förhåller sig till liknande SLR:er som utförts av andra forskare.

10.1.1 Litteraturgenomgångens styrkor och svagheter

Den litteraturgenomgång som presenterats i rapporten har en primär styrka: den har varit systematisk. Liknande genomgångar har gjorts av andra forskare men då oftast betydligt mindre systematiskt. Med tanke på hur viktiga systematiska metoder är inom vetenskapen kan denna genomgång rentav sägas vara mer vetenskaplig än flera andra genomgångar. En nyckel till att upprätthålla systematiken har varit att anamma de vägledningar som finns vad gäller utförandet av systematiska litteraturgenomgångar inom it-området. Delvis har andra systematiska litteraturgenomgångar använts som vägledningar men framförallt rör det sig om den uttryckliga vägledning som förmodligen är standardreferensen på området: Kitchenham och Charters (2007). Trots vägledningarna och denna rapports författares bästa ansträngningar att följa dem kan olika typer av metodfel uppstått. Några av de metodfel som insetts av författarna beskrivs nedan.

Det är alltid svårt att skapa en väl avvägd söksträng som inkluderar det som är relevant för forskningsfrågorna samtidigt som det som inte är relevant exkluderas. För att söksträngen skulle bli rätt utgick denna genomgång från söksträngarna i tidigare genomgångar. Därtill användes pilotsökning för att hitta rätt. Av dessa skäl har förmodligen söksträngen inga större brister men dess inkluderande av termen steganografi (ett ämne som senare insågs vara perifert) gjorde att mer manuell exkludering behövdes. I övrigt är det möjligt att relevanta artiklar missats och då särskilt när de inte fokuserar på detektion av illasinnad

nätverksbaserad exfiltrering. Deras fokus kan till exempel istället vara på exfiltreringens utförande där bara aspekter av detektionen nämns, eller på förbättring av nätverksstyrning med relevant men implicit säkerhetsvinkel.

Därtill är manuell exkludering av forskningsartiklarna inte möjligt att göra på ett helt systematiskt och objektivt sätt. Exempelvis var det ofta svårt att bedöma artiklarnas relevans utifrån deras sammanfattningar (eng. abstract), vilket är förståeligt med tanke på att artiklar inom it-området typiskt har undermåliga sammanfattningar (se Kitchenham och Brereton, 2013, sid. 26). Genom att i tveksamma fall läsa artiklarnas fulltexter och använda snöbollning har problemet minskats vad gäller exkludering men det kan ha fallit bort några artiklar som borde varit med. Det är också möjligt att exkluderingen borde varit hårdare och exempelvis exkluderat konferensbidrag för att fokusera på tidskriftsartiklar vilka normalt håller högre kvalitet. Ingen bedömning gjordes dock av huruvida det fanns genomgående skillnader mellan de två typerna av artiklar och därmed om resultatet hade blivit annorlunda om konferensbidragen exkluderats.

Vad gäller extraheringen av data från artiklarna och syntes av data var det svårt att från början avgöra vilka kategoriseringar som var lämpliga. Även här nyttjades pilottester för att sedan förfina kategoriseringarna. Därtill gjordes dubbelkontroller i form av att vissa artiklar extraherades av båda författarna av den här rapporten. Det är dock möjligt att resultatet ändå blivit färgat av subjektiva bedömningar. Inte minst är det möjligt att andra forskare skulle gjort andra bedömningar av både val av kategoriseringsmall och utförandet av kategoriseringen. Detta bedöms delvis i nästa avsnitt.

10.1.2 Jämförelse med andra litteraturgenomgångar

I bakgrundskapitlet beskrevs tre tidigare systematiska litteraturgenomgångar. En var Avila m.fl. (2021) som har ett annat fokus, i form av att detektera dataläckor med hjälp av olika typer av loggar. En annan var Sabir m.fl. (2022) som undersöker hur dataexfiltrering kan detekteras med hjälp av maskininlärning. Från den genomgången hämtades en del inspiration vad gäller kategorisering för extrahering och syntes. Överlappen i inkluderade artiklar var små med denna rapports genomgång. Det går dock att se mindre likheter i vilka exfiltreringsverktyg (för tunnling) som tas upp i den genomgången med de som tas upp i denna rapports genomgång. Därtill går det att jämföra vilka detektionsalgoritmer som är vanligast, justerat för antal artiklar. Jämfört med denna rapport ser då Sabir m.fl. (2022) ungefär lika ofta att beslutsträd, SVM och KNN används. Å andra sidan är djupinlärning något vanligare där medan Naive Bayes är betydligt vanligare i denna rapport. Med tanke på att Sabir m.fl. (2022) har ett större fokus på maskininlärning är det rimligt att det också blir något fler artiklar med fokus på mer avancerade maskininlärningsalgoritmer.

Ämnesmässigt ligger Ullah m.fl. (2018) närmast denna genomgång. Dock har Ullah m.fl. (2018) ett bredare fokus på dataexfiltrering i allmänhet och tar även

upp motmedel som inte enbart baseras på detektion. En skillnad är också tidsmässig i och med att den publicerades i januari för nästan fem år sedan och då efter en förmodligen ganska lång granskningsprocess. Därmed gör åldersskäl att den inte kan täcka in mer än maximalt en tredjedel av de artiklar som täcks in här (enligt Figur 3 i avsnitt 5.1).¹⁰ Därför är det svårt att göra jämförelser i syfte att säkerställa att samma resultat uppnåtts med (ungefär) samma metod. En relevant jämförelse är däremot att se om huvuddragen av resultaten bestått över tid eller om det finns några trender i de senaste årens forskningsartiklar. Även kategoriseringarna skiljer sig dock vilket gör de flesta jämförelser otympliga. En mindre jämförelse som kan göras är vad gäller forskningsartiklarnas utvärderingar av detektionsmetoderna. I denna rapports genomgång gjorde artiklarna sällan mer än knapphändiga utvärderingar av exempelvis hur mycket data som kan exfiltreras och hur snabbt detektion kan ske. Ullah m.fl. (2018) observerade på liknande sätt att ganska många studiers detektionsmetoder inte utvärderades alls.

10.2 Praktiska implikationer

Den akademiska forskningslitteraturen kan ses som teoretisk utan tydlig koppling till den verklighet där organisationer måste skydda sig mot exfiltrering. I de följande avsnitten beskrivs de praktiska implikationerna. Det gäller artiklarnas relation till militära miljöer, till riktiga angrepp (med dess data, hotaktörer, och verktyg), till detektionsprodukter, samt till säkerhetsfunktioner som kompletterar detektion.

En fullständig analys av dessa aspekter ligger dock utanför ramen för denna rapport. I detta avsnitt ges enbart en första bild. Det första att notera är att det är svårt att göra bedömningarna eftersom artiklarna ofta ger knapphändiga detaljer om vad exakt det är som studeras och under vilka omständigheter. Med tanke på att många av artiklarna ger en knapphändig beskrivning av sin miljö och av en tänkt hotaktör blir det svårt att veta hur nära riktiga angrepp man är och vilka typer av organisationer som forskningen så småningom kan gagna. Å andra sidan är det rimligt att forskningsartiklar ger mer otydliga beskrivningar varpå detaljerna utkristalliseras när en detektionsmetod går från labbstudie till produkt. Sådan produktifiering syns inte i huvudsakligen akademisk forskningslitteratur.

10.2.1 Militära miljöer

Forskningen fokuserar oftast på miljöer som utgörs av universitetsnätverk eller labbmiljöer på universitet. Detta kan jämföras med vad som rapporterats vara de mest utsatta totalt för både oavsiktliga och avsiktliga dataläckage av

¹⁰ I själva verket är överlappet betydligt mindre, vilket inte förvånar med tanke på att det förutom tid skiljer i ämnesfokus.

personuppgifter: finans-, tjänste-, teknik- och industrisektorerna (IBM, 2022). Om universitetsnätverk skiljer sig väsentligt från nätverken i de sektorerna (och allt annat är lika) kommer artiklarnas resultat inte ha någon större relevans i praktiken. Och om nämnda sektorer ligger långt ifrån universitetsnätverken ligger förmodligen militära nätverk ännu längre ifrån. Hur ett militärt sammanhang spelar roll tas upp i detta avsnitt.

Bland forskningsartiklarna har enbart en artikel en militär miljö. Det är Aurisch m.fl. (2021) vilken studerar taktiska ad-hoc-nätverk. Sådana nätverk skiljer sig mot typiska kontorsnätverk genom att nätverksnoderna rör på sig mycket och på ett oberäkneligt sätt samt ofta helt saknar anslutning till andra noder under vissa perioder. Även när noderna har anslutning sker det typiskt på ett decentraliserat sätt nod till nod utan någon central nod. Dessutom präglas noderna av låg bandbredd. Tillsammans gör dessa egenskaper det svårare att få till välinformerad detektion av exfiltrering. Det blir till exempel svårare att samla in stora mängder data för att ta beslut på. Därtill blir vissa vanliga kännetecken oanvändbara för detektion, vilket exempelvis gäller tiden mellan paket. Å andra sidan försvåras även exfiltreringen på grund av paketförseningar i ett sådant nätverk. Artikeln nämner att en specialversion av exfiltreringsverktyget DET krävdes för att kunna använda det i nätverket. Artikeln nämner också att detektionsmetoden var upplagd för att detektera små filer motsvarande storleken för geografisk positionsdata och för privata nycklar. Detta är typisk information som behöver hemlighållas i sådana nätverk medan data som kreditkortsnummer (som också studerats i artiklarna) är mindre relevant. En annan typ av information som ligger nära militär användning studerades i Bortolameotti m.fl. (2017) där data var i form av känsliga dokument som läckts av Wikileaks. Kopplat till konfidentialitet kan det också finnas särskilda begränsningar i hur militära system kan basera detektionen på datas innehåll med hänvisning till stark kryptering och hårda regler kring vem som får ges åtkomst till informationen.

Det finns också andra aspekter som är speciella i vissa militära sammanhang. Tvärtemot exemplet med mobila ad-hoc-nätverk kan fördröjningar i vissa situationer vara mycket allvarliga. Det kan till exempel ställa krav på att detektionsmetoderna är väldigt snabba. Därtill kan hotaktörerna vara särskilt allvarliga och resursstarka. I artiklarna anges sällan någon särskild tänkt hotaktör men Marchetti m.fl. (2016) är ett undantag. Där ses hotaktören som en (påhittad) APT som är svår-detekterad genom dess avancerade tekniska förmåga och lågintensiva exfiltrering. Även Das m.fl. (2017) utgår från en APT. Å andra sidan finns i vissa militära system redan (kostsamma) skydd i form av stark logisk eller fysisk separering gentemot omvärlden. Förutom speciella hotaktörer har militära sammanhang också speciella försvarare. I militära sammanhang är kulturen starkt hierarkiskt och säkerhetsmedveten. Det gör det troligare att det finns en intention att följa säkerhetsreglerna vilket är positivt för säkerheten. Å andra sidan agerar personalen också i stressigare situationer vilket gör det svårare att faktiskt

fullfölja intentionen. För nätverksbaserad exfiltrering har personalens beteende hur som helst mindre relevans än om det gällt fysisk exfiltrering.

10.2.2 Riktiga angrepp och exfiltreringsverktyg

Forskningsartiklarna beskriver sällan vad det är för information som (är tänkta att) exfiltreras. När det nämns är det kreditkortsuppgifter, inloggningsuppgifter och dokument. I riktiga angrepp är det som exfiltreras typiskt av olika filformat som används av ordbehandlare, kalkylprogram och dylikt (McAfee, 2018). Om detta gör att forskningen fokuserar alltför lite på dokument går inte att säga med tanke på att så få artiklar tar upp formatet på data som ska exfiltreras. Liksom exfiltrerade data är hotaktörer också knapphändigt beskrivna i artiklarna. Dock faller det utanför ramen för denna rapport att göra en analys av riktiga angripare.

Artiklarna döljer exfiltreringen främst genom att skicka data på ovanliga ställen i paket (t.ex. i pakethuvuden) eller med protokoll (t.ex. DNS) som normalt inte används för användares dataöverföring. Hur vanligt detta är i praktiken har inte identifierats inom ramen för denna rapport. Sett till vilka tekniker som ingår i ATT&CK-ramverket är bland annat exfiltrering via molnlagring populärt (se Mitre, 2022). Vad gäller andra sätt att dölja exfiltreringen verkar forskningens tekniker ungefär lika ofta vara krypterade som okrypterade. I praktiken menar McAfee (2018) att en tredjedel av exfiltreringen sker krypterat (medräknat fysisk exfiltrering). Även i ATT&CK finns det flest exempel för okrypterad exfiltrering (se subteknikerna i Mitre, 2021a).

Artiklarna väljer främst protokollet DNS för att exfiltrera data. Detta har en koppling till praktiken eftersom riktiga angripare ofta använder DNS-förfrågningar för exfiltrering och då domännamn av särskilt kort längd för att ha stor plats för de data som ska exfiltreras (Nadler m.fl., 2019). Däremot är det ovanligt att forskningen försöker detektera HTTP-trafik. I praktiken kan HTTP-baserad exfiltrering dock vara tämligen vanligt med tanke på att mer än 90 % av alla APT:er använder HTTP-baserad styrning (C2) (Wang m.fl., 2016) och att exfiltrering över styrkanal är en vanlig teknik i ATT&CK-ramverket¹¹. Därtill är HTTP det mest stödda protokollet hos C2-verktyg som används vid säkerhetstestning enligt webbplatsen C2 Matrix (2022).

Ganska många artiklar har studerat protokollkombinationen DNS över HTTPS (DoH) de senare åren. I praktiken har DoH hittills använts för att etablera styrkanaler. Det gäller åtminstone för de båda skadliga koderna Godlua (Turing och Ye, 2019) och Sandworms Cyclops Blink (Hacquebord m.fl., 2022). Det finns också säkerhetstestningsverktyg med stöd för att genomföra DoH (se

¹¹ Se Mitre (2021b). Förutom att det finns en egen teknik för detta är antalet exempel för tekniken stort i jämförelse med antalen för övriga exfiltreringstekniker i ramverket. Antalen exempel är dock inte nödvändigtvis representativa.

Github 2021; 2022c). Däremot verkar inte DoH användas för exfiltrering i praktiken. Åtminstone nämns inga sådana fall i varken forskningsartiklarna eller de praktiska attacktekniker som tas upp i ramverket ATT&CK¹².

Att det finns gott om färdiga verktyg för att utföra exfiltrering märks tydligt i artiklarna då verktygen ofta används vid utvärdering av detektionsmetoderna. Verktygen används dock inte nödvändigtvis av riktiga angripare utan kan vara mest för forskning och säkerhetstester.

10.2.3 Produkter för detektion av dataexfiltrering

En utförlig studie av detektionsprodukter ligger utanför denna rapports ramar. Det finns till exempel inte utrymme att utvärdera produkter utifrån de många mått som används för utvärdering av forskningens detektionsmetoder. Vad som kan sägas är att jämfört med forskningen baseras produkternas detektion mer på vad det är för innehåll som exfiltreras än på abstrakta aspekter som kontexten. Exempelvis fokuserar djup paketinspektion (*DPI*) på innehåll och är en vanlig metod vid organisationsnätverkets utkant (Shabtai m.fl., 2012). Därtill finns skydd mot dataläckage (eng. *Data Loss Prevention*, *DLP*) vilka ofta kombinerar detektion av data som rör sig i nätverket med detektion av data som hanteras felaktigt i individuella ändpunkter (t.ex. datorer och mobiltelefoner). Generellt finns också olika typer av loggverktyg som går under förkortningar som *SIEM* och *SOAR*. Därtill finns några år tillbaka en ny sorts produkt som sägs vara till för utökad detektion och respons (eng. *Extended Detection and Response*, *XDR*).

Produkters detektion verkar oftare vara regelbaserad än anomalibaserad eller maskininlärningsbaserad. Det vore också logiskt om de flesta produkter satsade på enklast möjliga detektion. Å andra sidan visade en (icke-vetenskaplig) undersökning av IBM (2022) att införandet av en AI-plattform är det bästa skyddet mot dataläckage av personuppgifter.

En speciell sak med detektionsprodukter är också att de ofta riktar sig mot både illasinnad exfiltrering och oavsiktlig utförsel av data. Hur vanligt det är att forskningsartiklarna har båda dessa fokus har inte utretts systematiskt i genomgången. Men ganska många av de manuellt exkluderade artiklarna exkluderades för att de fokuserade på misstag snarare än avsiktlig illasinnad exfiltrering.

10.2.4 Komplement till detektion

Detektion av illasinnad nätverksbaserad exfiltrering är inte det enda sättet att försöka skydda sig mot exfiltreringen. Det är inte heller tillräckligt att detektera utan att också hantera den kunskap som detektionen ger. Den tidigare

¹² Det saknas nämligen exempel på DoH för exfiltrering i Mitre (2020b).

litteraturgenomgången Ullah m.fl. (2018) ger utöver detektion även alternativen prevention (i förväg) och utredning (efteråt). Som prevention ser artikeln fem varianter: att fylla databas och nätverk med oviktiga data för att angriparen inte ska hitta det viktiga; att inte lagra data på en central plats; att kryptera data; att införa åtkomstkontroll; samt att genomföra informations säkerhetsklassning. Som utredning ser artikeln tre varianter: vattenstämplar; forensik; samt att spåra upp angriparen. Utöver artikelns förslag finns åtminstone två ytterligare alternativ: hotjakt och avskräckning. Hotjakt går ut på att leta efter spår från hotaktörens tidiga agerande och kan avslöja denne innan exfiltreringen påbörjats. Avskräckning går ut på att påverka hotaktörens intention genom att göra denne medveten om försvararens möjligheter att hitta och bekämpa hotaktören.

10.3 Framtida forskning som behövs

Mycket forskning existerar redan men det finns också behov av mer forskning för att på sikt möjliggöra bättre försvarsmekanismer. Bland forskningen som finns är den för det mesta utförd med begränsade metoder. Framtida forskning behöver exempelvis gå från att så starkt fokusera på universitetsnät och vanliga kontorsnät till att fokusera mer på mjukvarudefinierade nätverk, sakernas internet, industriella kontrollsystem och militära miljöer. Därtill behövs mer detaljerade angreppsscenarioer, i stort sett oavsett vilket scenario som väljs. Dessutom behöver både data och miljöer vara större för att möjliggöra mer realistiska experiment. Vissa forskare har redan inlett samarbeten genom att använda varandras verktyg och dataset. Sådana samarbeten behöver fortsätta och utökas.

Forskningen behöver som minst hålla jämna steg med hotaktörerna när de kommer på nya knep. Exempelvis behövs forskning om hur detektionsmetoderna ska hantera att hotaktörerna minskar sina spår med trafikomvandling. Sådan forskning har bland annat inletts i Xu m.fl. (2022) samt Sadeghzadeh m.fl. (2021). Därtill behöver forskning för att kunna detektera exfiltrering som sker via videomöten, exempelvis med den teknik som föreslås av Barradas m.fl. (2017) samt via blockkedjeteknik, exempelvis i form av de tekniker som sammanställts av Tian m.fl. (2020). Mer forskning behövs också för andra tekniker som blir populärare. Som genomgången visat har exempelvis forskningen de senaste åren börjat undersöka detektion för DNS över HTTPS samt för IPv6 men det finns goda möjligheter till mer forskning där. Det behövs också mer forskning om hur exfiltrering ska detekteras i miljöer som inte är typiska kontorsnätverk utan är mjukvarudefinierade nätverk, sakernas internet industriella kontrollsystem eller militära miljöer.

Förutom att ta fram en ny detektionsmetod för varje ny exfiltreringsteknik kan forskare också ha mer framförhållning. Det vore därför bra om detektionsmetoderna utvecklades med ändringsbarhet och anpassningsbarhet i åtanke. Detektionsmetoder kan också vara mer generella och därmed direkt

fungera mot en stor mängd exfiltreringstekniker. Exempelvis nämner Caviglione (2021) kandidater för mer hotoberoende detektion. Kandidatmetoderna är främst inriktade på detektion på en enskild maskin och baseras på samband mellan processers aktivitet samt på analys av datorns energikonsumtion. Ett annat sätt att få fram mer generella detektionsmetoder är att forskarna inte själva väljer vad metoden ska basera sin detektion på, utan att en maskininlärningsalgoritm istället själv får välja vilka kännetecken den ska leta efter. Sådan automatik har framförallt använts i genomgångens artiklar som nyttjar djupinlärningsalgoritmer. Ett ytterligare sätt att ta fram mer generella detektionsmetoder är att se till att tränings- och testdata är mer generella. På så vis skulle metoderna inte bara se bra ut på papper utan också ha större chans att vara giltiga utanför labbet. Att exempelvis i labbmiljö detektera baserat på antalet gånger "confidential" nämns i en text är undermåligt om metoden ska användas för att detektera ett "konfidentiellt" svenskspråkigt dokument. På motsvarande sätt bör nog inte samma tyngd läggas på teckenentropi i ett språk med färre bokstäver (som engelska) än ett med fler (som svenska). En mer generell metod skulle tillåta sådana variationer i användningen liksom ha inställningar för hur viktigt det är att stoppa ett visst hot i en viss miljö samt vilka negativa konsekvenser för legitim användning som kan tolereras.

Forskning bör samtidigt reflektera över att nya detektionsmetoder av exfiltrering inte bara kan användas för ändamål som forskarna anser vara illasinnade utan också för andra ändamål såsom totalitära staters censur och övervakning av sina medborgare. Att detektionsmetoder kan användas på dessa sätt har lyfts av Wendzel m.fl. (2015).

11 Slutsatser

I detta kapitel beskrivs slutsatser i form av svaren på forskningsfrågorna, vad rapportens bidrag är gentemot annan forskning, vilken ytterligare forskning som behövs, samt i allmänhet vad läsaren bör ta med sig från rapporten.

11.1 Vilka metoder för detektion av illasinnad nätverksbaserad exfiltrering av data från it-system studeras i forskningslitteraturen?

Rubriken utgör också rapportens övergripande forskningsfråga. Svaret på denna forskningsfråga behöver förstås och ges i ett sammanhang. Sammanhanget utgörs av vad det är för typ av forskning, med dess ursprung och kvalitet samt vilka miljöer, angreppsscenarier och exfiltreringstekniker som forskningen om detektion studerar. I rapportens inledning formulerades därför forskningsfrågor även om dessa aspekter. Svaren på forskningsfrågorna ges i de följande avsnitten. På detta sätt besvaras också den övergripande forskningsfrågan.

11.1.1 Vad är studiernas ursprung och hur är deras kvalitet?

Litteraturgenomgången inkluderar 48 relevanta artiklar utgivna över drygt tio år. Av dessa artiklar är knappt fyra av tio tidskriftsartiklar medan drygt sex av tio utgör konferensbidrag.

Artiklarnas syften är tydliga i termer av valt ämne och varför forskningen görs. Vad gäller mer specifika mål är det däremot ofta otydligt beskrivet på så vis att experiment saknar tydliga forskningsfrågor eller hypoteser. Därtill är artiklarnas metodbeskrivningar för det mesta begränsat beskrivna, vilket gör det svårt eller omöjligt att förstå forskningen fullt ut. Ingen artikel är dock varken usel eller perfekt när det gäller metodbeskrivningarna.

11.1.2 Vilken miljö antar studierna att exfiltreringen sker i?

För att kunna studera exfiltrering bygger de flesta artiklarna upp en miljö med legitima data vilka ofta samlats in från riktiga nätverk. I andra fall har inte data samlats in utan tas fram på syntetisk väg i labbmiljö. En annan möjlighet är att använda riktiga data som används live i riktiga nätverk men det används sällan. De flesta artiklar tar fram sina egna legitima data men det är också vanligt att använda dataset som sammanställts av andra forskare eller organisationer.

De nätverk som studeras utgörs oftast av universitetsnätverk eller labbmiljöer på universitet. Enbart en artikel har en uttrycklig militär nätverksmiljö. Det gäller ett taktiskt ad-hoc-nätverk där noder har oberäknliga anslutningar vilket gör att exfiltreringstekniker måste anpassas och detektionsmetoder vara mer decentraliserade.

Vad gäller storleken på artiklarnas nätverk beskrivs den sällan och för de artiklar där det beskrivs finns det stora variationer från en användare till 100 000 eller mer samt från korta studier på timmar till långa studier på månader.

11.1.3 Vilket angreppsscenario antar studierna att exfiltreringen sker i?

Forskningsartiklarna beskriver sällan vilken information som är tänkta att exfiltreras. När den nämns utgörs data av kreditkortsuppgifter, inloggningsuppgifter och dokument.

Hotaktörerna är också sällan beskrivna. När de nämns är de ibland insiders, ibland outsiders. Vad gäller hotaktörernas kapacitet beskrivs den i ett fåtal artiklar utifrån att hotaktörerna är APT:er, det vill säga avancerade och ihärdiga hotaktörer som dessutom ofta har statsstöd. Sådana hotaktörer är också särskilt relevanta i ett militärt sammanhang.

11.1.4 Vilka exfiltreringstekniker försöker studierna detektera?

Artiklarna försöker detektera exfiltreringstekniker där data främst döljs genom att de skickas på ovanliga ställen i paket (exempelvis i pakethuvuden) eller med protokoll (såsom DNS) som normalt inte används för användares dataöverföring. Det finns också andra sätt för döljande varav bland annat kryptering men det rör sig ungefär lika ofta om okrypterad exfiltrering. Vad gäller protokoll baseras teknikerna oftast på DNS samt DNS i en relativt ny protokollkombination med HTTPS. Däremot är det ovanligt att forskningen använder HTTP vilket däremot verkar vanligt i riktiga angrepp. För att implementera exfiltreringsteknikerna använder artiklarna ofta färdiga verktyg såsom Iodine, Dnscat2, Dns2tcp och DET.

11.1.5 Hur fungerar studiernas detektionsmetoder?

Artiklarnas detektionsmetoder försöker oftast detektera exfiltrering genom skillnader i entropi, såsom teckenfrekvenser i domännamn. Även vanligt är att basera detektionen på tidsaspekter (såsom tiden mellan paket), stränglängder (exempelvis hos domännamn), trafikflöden (till exempel antal förfrågningar per domän) samt pakethuvudinnehåll i form av värden hos specifika fält.

Vad gäller detektionsmetodernas algoritmer rör det sig oftast om djupinlärningsbaserade sådana och i andra hand traditionell maskininlärning såsom beslutsträd och SVM. Däremot är det sällan som algoritmerna är signaturbaserade med exempelvis mönsterigenkänning stödd av reguljära uttryck.

Hur detektionsmetoderna fungerar kan också skönjas utifrån artiklarnas utvärderingar av metoderna. I utvärderingarna används många olika mått. I grunden rör det sig om att undersöka hur ofta detektionen detekterar när den borde detektera, inte detekterar när den inte borde detektera, samt tvärtom när den detekterar trots att den inte borde, och när den inte detekterar trots att den borde. Dessa grunder kombineras till ett antal olika mått och det finns ingen standardisering för vilka mått som används. Därtill är utvärderingarna begränsade, exempelvis i form av att detektionsmetodernas snabbhet sällan anges ordentligt.

11.2 Rapportens bidrag

Den här rapporten har genomfört en systematisk litteraturgenomgång för att besvara forskningsfrågor om detektion av illasinnad nätverksbaserad dataexfiltrering. Liknande genomgångar har gjorts av andra forskare men då oftast betydligt mindre systematiskt. Med tanke på hur viktiga systematiska metoder är inom vetenskapen kan denna genomgång rentav sägas vara mer vetenskaplig än flera andra genomgångar. Det finns dock även andra systematiska litteraturgenomgångar vilka kan jämföras med denna rapport. En är Sabir m.fl. (2022) vilken har ett snävare fokus på maskininlärning än denna rapport har. En annan är Ullah m.fl. (2018) vilken har ett ämnesfokus som mer liknar denna rapport men som är fem år äldre och därmed saknar en stor del av den forskning som ingår i denna rapports genomgång.

11.3 Framtida forskning

Utöver de forskningsartiklar som ingår i denna rapports genomgång finns det gott om utrymme för mer forskning. En aspekt är att vidga miljöerna för exfiltreringen från universitetsnätverk och vanliga kontorsnät till mjukvarudefinierade nätverk, sakernas internet, industriella kontrollsystem och militära miljöer. Därtill behövs mer detaljerade angreppscenarier, i stort sett oavsett vilket scenario som väljs. Dessutom behöver både data och miljöer vara större för att möjliggöra mer realistiska experiment. Vissa forskare har redan inlett samarbeten genom att använda varandras verktyg och dataset. Sådana samarbeten behöver fortsätta och utökas.

Vad gäller exfiltreringstekniker gäller det i första hand att hålla jämna steg med angripare som hittar nya vägar. Exempelvis behövs forskning om hur detektionsmetoderna ska hantera att hotaktörerna minskar sina spår med

trafikomvandling. Därtill behövs mer forskning om detektion av exfiltrering som sker via videomöten, blockkedjeteknik, DNS över HTTPS, IPv6 och andra nyare protokoll. Därtill vore det bra om detektionsmetoderna blev mer generella genom att exempelvis satsa mer på algoritmer som är mer autonoma och på dataset som är mer allmängiltiga. Samtidigt behöver forskningen väga in risken att nya detektionsmetoder också kan användas till att upprätthålla censur och begränsningar av dissidenter i mindre demokratiska regimer.

11.4 Avslutning

I en rapport om akademisk forskningslitteratur kan det som avslutning vara på sin plats med några ord som balanserar forskningens ambitioner med realism och vakenhet. En angripare kan nyttja ett stort antal olika tekniker för att exfiltrera data och ingen universallösning finns för att detektera alla tekniker. Vilka detektionsmetoder som är bäst för en försvarare att nyttja beror på det sammanhang där angreppet sker. Det är också viktigt att inse att det inte går att stoppa alla möjligheter till illasinnad exfiltrering. Om ett nätverk ska vara användbart måste det nämligen tillåta överföring och sådan överföring ger nästan alltid en chans för angripare. En asymmetri finns där angripare bara behöver hitta en väg ut medan försvarare måste hålla ögonen öppna för alla sätt att exfiltrera data.

12 Referenser

Referenserna är uppdelade i allmänna referenser (för hela rapporten), relaterade FOI-rapporter (för bakgrundskapet), genomgångar (för bakgrundskapet) och i identifierade forskningsartiklar (för primärkällorna som ingår i den systematiska litteraturgenomgången).

12.1 Allmänna referenser

- Antwerp R.C.V. 2011. Exfiltration Techniques: An Examination And Emulation. Master's thesis, University of Delaware.
- Barradas D., Santos N., Rodrigues L.E. 2017. DeltaShaper: Enabling Unobservable Censorship-resistant TCP Tunneling over Videoconferencing Streams. Proc. Priv. Enhancing Technol., 2017(4), 5-22.
<https://doi.org/10.1515/popets-2017-0024>
- BBC. 2018. British Airways boss apologises for 'malicious' data breach.
<https://www.bbc.com/news/uk-england-london-45440850>
- BBC. 2020. British Airways fined £20m over data breach.
<https://www.bbc.com/news/technology-54568784>
- C2 Matrix. 2022. <https://www.thec2matrix.com/matrix> (Besökt 28 nov 2022.)
- Caviglione L. 2021. Trends and challenges in network covert channels countermeasures. Applied Sciences, 11(4), 1641.
<https://doi.org/10.3390/app11041641>
- Github. 2017. Dns2tcp. <https://github.com/alex-sector/dns2tcp> (Besökt 28 nov 2022.)
- Github. 2019. DET. <https://github.com/PaulSec/DET> (Besökt 28 nov 2022.)
- Github. 2021. Deimosc2. <https://github.com/DeimosC2/DeimosC2> (Besökt 28 nov 2022.)
- Github. 2022a. Iodine. <https://github.com/yarrick/iodine> (Besökt 28 nov 2022.)
- Github. 2022b. Dnscat2. <https://github.com/iagox86/dnscat2> (Besökt 28 nov 2022.)
- Github. 2022c. Godoh. <https://github.com/sensepost/goDoH> (Besökt 28 nov 2022.)
- Hacquebord F., Hilt S., Merces F. 2022. Cyclops Blink Sets Sights on Asus Routers. Trend Micro.
https://www.trendmicro.com/en_us/research/22/c/cyclops-blink-sets-sights-on-asus-routers--.html

- IBM. 2022. Cost of a Data Breach Report 2022. IBM Security.
- ICO. 2018. Penalty Notice. Section 155, Data protection Act 2018. Case ref: COM0804337. <https://ico.org.uk/media/action-weve-taken/mpns/2618524/marriott-international-inc-mpn-20201030.pdf>
- Jacobs L. 2018. Waiting for goDoH. Sensepost. <https://sensepost.com/blog/2018/waiting-for-godoh/>
- Kitchenham B., Brereton P. 2013. A systematic review of systematic review process research in software engineering. Information and software technology, 55(12), 2049-2075. <https://doi.org/10.1016/j.infsof.2013.07.010>
- Kitchenham B., Charters S. 2007. Guidelines for performing systematic literature reviews in software engineering. Version 2.3. EBSE Technical Report EBSE-2007-01. Keele University & University of Durham.
- Leopold J. 2019. Here Are The Never-Before-Seen US Government Damage Reports Made In The WikiLeaks Aftermath. BuzzFeed News. <https://www.buzzfeednews.com/article/jasonleopold/here-are-the-never-before-seen-us-government-damage-reports>
- Mcafee. 2018. Grand Theft Data. Data exfiltration study: Actors, tactics, and detection.
- Mitre. 2020a. ATT&CK. Proxy: Domain Fronting. Enterprise v12. ID: T1090.004. (Senast ändrad 16 sep 2020.) <https://attack.mitre.org/versions/v12/techniques/T1090/004/>
- Mitre. 2020b. ATT&CK. Protocol Tunneling. Enterprise v12. ID: T1572. (Senast ändrad 27 mars 2020.) <https://attack.mitre.org/versions/v12/techniques/T1572/>
- Mitre. 2021a. ATT&CK. Exfiltration Over Alternative Protocol. Enterprise v12. ID: T1048. (Senast ändrad 15 oktober 2021.) <https://attack.mitre.org/versions/v12/techniques/T1048/>
- Mitre. 2021b. ATT&CK. Exfiltration Over C2 Channel. Enterprise v12. ID: T1041. (Senast ändrad 15 oktober 2021.) <https://attack.mitre.org/versions/v12/techniques/T1041/>
- Mitre. 2022. ATT&CK. Exfiltration Over Web Service: Exfiltration to Cloud Storage. Enterprise v12. ID: T1567.002. (Senast ändrad 30 aug 2022.) <https://attack.mitre.org/versions/v12/techniques/T1567/002/>
- Nist. 2020. NIST Special Publication 800-53. Revision 5. Security and Privacy Controls for Information Systems and Organizations.
- Packetlabs. 2018. Case Study: Marriott Data Breach. <https://www.packetlabs.net/posts/case-study-marriott-data-breach/>

- Ring M., Wunderlich S., Scheuring D., Landes D., Hotho A. 2019. A survey of network-based intrusion detection data sets. *Computers & Security*, 86, 147-167. <https://doi.org/10.1016/j.cose.2019.06.005>
- Sadeghzadeh A.M., Shiravi S., Jalili R. 2021. Adversarial network traffic: Towards evaluating the robustness of deep-learning-based network traffic classification. *IEEE Transactions on Network and Service Management*, 18(2), 1962-1976. <https://doi.org/10.1109/TNSM.2021.3052888>
- Sudo T., Sato T. 2020. Mitsubishi Electric attack likely stole data on new missile. *Asahi Shimbun*. <https://www.asahi.com/ajw/articles/13388776>
- Tian J., Xiong G., Li Z., Guo G. 2020. A Survey of Key Technologies for Constructing Network Covert Channel. *Security and Communication Networks Volume 2020*, Article ID 8892896, <https://doi.org/10.1155/2020/8892896>
- Turing A., Ye G. 2019. An Analysis of Godlua Backdoor. *Netlab*. <https://blog.netlab.360.com/an-analysis-of-godlua-backdoor-en/>
- Ullah F., Edwards M., Ramdhany R., Chitchyan R., Babar M.A., Rashid A. 2018. Data exfiltration: A review of external attack vectors and countermeasures. *Journal of Network and Computer Applications*, 101, 18-54. <https://doi.org/10.1016/j.jnca.2017.10.016>
- Wang X., Zheng K., Niu X., Wu B., Wu C. 2016. Detection of Command and Control in Advanced Persistent Threat based on Independent Access. *Communication and Information Systems Security Symposium*. <https://doi.org/10.1109/ICC.2016.7511197>
- Wang Y., Zhou A., Liao S., Zheng R., Hu R., Zhang L. 2021. A comprehensive survey on DNS tunnel detection. *Computer Networks*. <https://doi.org/10.1016/j.comnet.2021.108322>
- Wendzel S., Zander S., Fechner B., Herdin C. 2015. Pattern-based survey and categorization of network covert channel techniques. *ACM Computing Surveys (CSUR)*, 47(3), 1-26. <https://doi.org/10.1145/2684195>
- Xu Z., Khan H., Muresan R. 2022. TMorph: A Traffic Morphing Framework to Test Network Defenses Against Adversarial Attacks. In *2022 International Conference on Information Networking (ICOIN)* (pp. 18-23). IEEE. <https://doi.org/10.1109/ICOIN53446.2022.9687165>

12.2 Relaterade FOI-rapporter

- Karlzén H., Sommestad T. 2022. Indicators of Compromise Suitable for Use in Threat Hunting Training. FOI Memo 7788. Totalförsvarets forskningsinstitut.

Lundholm K., Karlzén H., Löfvenberg J. 2010. Dolda Kanaler. En översikt. FOI-R--3037--SE. Totalförsvarets forskningsinstitut.

Stenborg K.-G., Allvar J., Haapalahti G., Herberthson M. 2010. Vattenmärkning av sensordata. Slutrapport. FOI-R--3138--SE. Totalförsvarets forskningsinstitut.

12.3 Genomgångar

Akram A., Mushtaq M., Bhatti M.K., Lapotre V., Gogniat G. 2020. Meet the Sherlock Holmes' of side channel leakage: A survey of cache SCA detection techniques. *IEEE Access*, 8, 70836-70860.
<https://doi.org/10.1109/ACCESS.2020.2980522>

Alneyadi S., Sithirasenan E., Muthukkumarasamy V. 2016. A survey on data leakage prevention systems. *Journal of Network and Computer Applications*, 62, 137-152. <https://doi.org/10.1016/j.jnca.2016.01.008>

Ávila R., Khoury R., Khoury R., Petrillo F. 2021. Use of security logs for data leak detection: a systematic literature review. *Security and communication networks*. <https://doi.org/10.1155/2021/6615899>

Caviglione L. 2021. Trends and challenges in network covert channels countermeasures. *Applied Sciences*, 11(4), 1641.
<https://doi.org/10.3390/app11041641>

Cheng L., Liu F., Yao D. 2017. Enterprise data breach: causes, challenges, prevention, and future directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(5), e1211.
<https://doi.org/10.1002/widm.1211>

Conti M., Li Q.Q., Maragno A., Spolaor R. 2018. The dark side (-channel) of mobile devices: A survey on network traffic analysis. *IEEE communications surveys & tutorials*, 20(4), 2658-2713.
<https://doi.org/10.1109/COMST.2018.2843533>

Sabir B., Ullah F., Babar M A., Gaire R. 2022. Machine learning for detecting data exfiltration: a review. *ACM Computing Surveys (CSUR)*, 54(3), 1-47.
<https://doi.org/10.1145/3442181>

Shabtai A., Elovici Y., Rokach L. 2012. A survey of data leakage detection and prevention solutions. *Springer Science & Business Media*.
<https://doi.org/10.1007/978-1-4614-2053-8>

Tian J., Xiong G., Li Z., Guo G. 2020. A Survey of Key Technologies for Constructing Network Covert Channel. *Security and Communication Networks Volume 2020*, Article ID 8892896.
<https://doi.org/10.1155/2020/8892896>

- Ullah F., Edwards M., Ramdhany R., Chitchyan R., Babar M. A., Rashid A. 2018. Data exfiltration: A review of external attack vectors and countermeasures. *Journal of Network and Computer Applications*, 101, 18-54. <https://doi.org/10.1016/j.jnca.2017.10.016>
- Wang Y., Zhou A., Liao S., Zheng R., Hu R., Zhang L. 2021. A comprehensive survey on DNS tunnel detection. *Computer Networks*. <https://doi.org/10.1016/j.comnet.2021.108322>
- Wendzel S., Zander S., Fechner B., Herdin C. 2015. Pattern-based survey and categorization of network covert channel techniques. *ACM Computing Surveys (CSUR)*, 47(3), 1-26. <https://doi.org/10.1145/2684195>

12.4 Identifierade forskningsartiklar

- Ahmed J., Habibi Gharakheili H., Raza Q., Russell C., Sivaraman V. 2020. Monitoring Enterprise DNS Queries for Detecting Data Exfiltration from Internal Hosts. *IEEE Transactions on Network and Service Management*. <https://doi.org/10.1109/TNSM.2019.2940735>
- Aiello M., Mongelli M., Cambiaso E., Papaleo G. 2016. Profiling DNS tunneling attacks with PCA and mutual information. *Logic Journal of the IGPL*. <https://doi.org/10.1093/jigpal/jzw056>
- Al-Bataineh A., White G. 2012. Analysis and detection of malicious data exfiltration in web traffic. *Proceedings of the 2012 7th International Conference on Malicious and Unwanted Software, Malware 2012*. <https://doi.org/10.1109/MALWARE.2012.6461004>
- Al-Eidi S., Darwish O., Chen Y., Husari G. 2021. SnapCatch: Automatic Detection of Covert Timing Channels Using Image Processing and Machine Learning. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2020.3046234>
- Almusawi A., Amintoosi H. 2018. DNS tunneling detection method based on multilabel support vector machine. *Security and Communication Networks*. <https://doi.org/10.1155/2018/6137098>
- Altuncu M.A., Gülağiz F.K., Özcan H., Bayir Ö.F., Gezgin A., Niyazov A., Çavuşlu M.A., Şahin S. 2021. Deep Learning Based DNS Tunneling Detection and Blocking System. *Advances in Electrical and Computer Engineering*. <https://doi.org/10.4316/AECE.2021.03005>
- Aurisch T., Chacon P.C., Jacke A. 2021. Mobile cyber defense agents for low throughput DNS-based data exfiltration detection in military networks. 2021 *International Conference on Military Communication and Information Systems, ICMCIS 2021*. <https://doi.org/10.1109/ICMCIS52405.2021.9486400>

- Ayub M.A., Smith S., Siraj A. 2019. A protocol independent approach in network covert channel detection. Proceedings - 22nd IEEE International Conference on Computational Science and Engineering and 17th IEEE International Conference on Embedded and Ubiquitous Computing, CSE/EUC 2019. <https://doi.org/10.1109/CSE/EUC.2019.00040>
- Bortolameotti R., Van Ede T., Caselli M., Everts M.H., Hartel P., Hofstede R., Jonker W., Peter A. 2017. DECANTeR: DEtECTION of Anomalous outbouNd HTTP Traffic by Passive Application Fingerprinting. ACM International Conference Proceeding Series. <https://doi.org/10.1145/3134600.3134605>
- Bubnov Y. 2018. DNS Tunneling Detection Using Feedforward Neural Network. European Journal of Engineering Research and Science. <https://doi.org/10.24018/ejeng.2018.3.11.963>
- Buczak A.L., Hanke P.A., Cancro G.J., Toma M.K., Watkins L.A., Chavis J.S. 2016. Detection of tunnels in PCAP data by random forests. Proceedings of the 11th Annual Cyber and Information Security Research Conference, CISRC 2016. <https://doi.org/10.1145/2897795.2897804>
- Cabaj K., Mazurczyk W., Nowakowski P., Zórawski P. 2018. Towards distributed network covert channels detection using data mining-based approach. ACM International Conference Proceeding Series. <https://doi.org/10.1145/3230833.3233264>
- Chen S., Lang B., Liu H., Li D., Gao C. 2021. DNS covert channel detection method using the LSTM model. Computers and Security. <https://doi.org/10.1016/j.cose.2020.102095>
- Chiu C.-K., Chang H.-H., Mao C.-H., Wei T.-E. 2019. Counterfeit Fingerprint Detection of Outbound HTTP Traffic with Graph Edit Distance. DSC 2018 - 2018 IEEE Conference on Dependable and Secure Computing. <https://doi.org/10.1109/DESEC.2018.8625111>
- Chowdhary A., Bhowmik M., Rudra B. 2021. DNS tunneling detection using machine learning and cache miss properties. Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021. <https://doi.org/10.1109/ICICCS51141.2021.9432279>
- Chung M.-H., Chignell M., Wang L., Jovicic A., Raman A. 2020. Interactive Machine Learning for Data Exfiltration Detection: Active Learning with Human Expertise. Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics. <https://doi.org/10.1109/SMC42975.2020.9282831>
- Cui Q., Jourdan G.-V., Bochmann G.V., Onut I.-V. 2021. Proactive Detection of Phishing Kit Traffic. Lecture Notes in Computer Science (including

- subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-030-78375-4_11
- D'Angelo G., Castiglione A., Palmieri F. 2022. DNS tunnels detection via DNS-images. Information Processing and Management. <https://doi.org/10.1016/j.ipm.2022.102930>
- Das A., Shen M.-Y., Shashanka M., Wang J. 2017. Detection of exfiltration and tunneling over DNS. Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017. <https://doi.org/10.1109/ICMLA.2017.00-71>
- Davis J.J., Foo E. 2016. Automated feature engineering for HTTP tunnel detection. Computers and Security. <https://doi.org/10.1016/j.cose.2016.01.006>
- Dijk A. 2021. Detection of Advanced Persistent Threats using Artificial Intelligence for Deep Packet Inspection. Proceedings - 2021 IEEE International Conference on Big Data, Big Data 2021. <https://doi.org/10.1109/BigData52589.2021.9671464>
- Ding S., Zhang D., Ge J., Yuan X., Du X. 2021. Encrypt DNS Traffic: Automated Feature Learning Method for Detecting DNS Tunnels. 2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom. <https://doi.org/10.1109/ISPA-BDCloud-SocialCom-SustainCom52081.2021.00056>
- Do V.T., Engelstad P., Feng B., van Do T. 2017. Detection of DNS tunneling in mobile networks using machine learning. Lecture Notes in Electrical Engineering. https://doi.org/10.1007/978-981-10-4154-9_26
- Durkota K., Lisý V., Kiekintveld C., Horák K., Božanský B., Pevný T. 2017. Optimal Strategies for Detecting Data Exfiltration by Internal and External Attackers. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-319-68711-7_10
- Ellens W., Zuraniewski P.Ě., Sperotto A., Mandjes M., Meeuwissen E. 2013. Flow-Based Detection of DNS Tunnels. IFIP International Conference on Autonomous Infrastructure, Management and Security. https://doi.org/10.1007/978-3-642-38998-6_16
- Eremeev M.A., Nefedov V.S., Ostrovskii A.S., Semchenkov D.A. 2021. The Use of Beacon Signals to Detect Covert Channels in DNS Traffic. Automatic Control and Computer Sciences. <https://doi.org/10.3103/S0146411621080095>

- He G., Zhang T., Ma Y., Xu B. 2015. A Novel Method to Detect Encrypted Data Exfiltration. Proceedings - 2014 2nd International Conference on Advanced Cloud and Big Data, CBD 2014. <https://doi.org/10.1109/CBD.2014.40>
- Iglesias F., Annessi R., Zseby T. 2016. DAT detectors: uncovering TCP/IP covert channels by descriptive analytics. Security and Communication Networks. <https://doi.org/10.1002/sec.1531>
- Ishikura N., Kondo D., Vassiliades V., Iordanov I., Tode H. 2021. DNS Tunneling Detection by Cache-Property-Aware Features. IEEE Transactions on Network and Service Management. <https://doi.org/10.1109/TNSM.2021.3078428>
- King J., Bendiab G., Savage N., Shiaeles S. 2021. Data exfiltration: Methods and detection countermeasures. Proceedings of the 2021 IEEE International Conference on Cyber Security and Resilience, CSR 2021. <https://doi.org/10.1109/CSR51186.2021.9527962>
- Lambion D., Josten M., Olumofin F., De Cock M. 2020. Malicious DNS Tunneling Detection in Real-Traffic DNS Data. Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020. <https://doi.org/10.1109/BigData50022.2020.9378418>
- Lin H., Liu G., Yan Z. 2019. Detection of Application-Layer Tunnels with Rules and Machine Learning. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-030-24907-6_33
- Marchetti M., Pierazzi F., Colajanni M., Guido A. 2016. Analysis of high volumes of network traffic for Advanced Persistent Threat detection. Computer Networks. <https://doi.org/10.1016/j.comnet.2016.05.018>
- Mitsuhashi R., Satoh A., Jin Y., Iida K., Shinagawa T., Takai Y. 2021. Identifying Malicious DNS Tunnel Tools from DoH Traffic Using Hierarchical Machine Learning Classification. International Conference on Information Security. https://doi.org/10.1007/978-3-030-91356-4_13
- Montazerishatoori M., Davidson L., Kaur G., Habibi Lashkari A. 2020. Detection of DoH Tunnels using Time-series Classification of Encrypted Traffic. 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/CyberSciTech). <https://doi.org/10.1109/DASC-PICom-CBDCCom-CyberSciTech49142.2020.00026>

- Nadler A., Aminov A., Shabtai A. 2019. Detection of malicious and low throughput data exfiltration over the DNS protocol. Computers and Security. <https://doi.org/10.1016/j.cose.2018.09.006>
- Nafea H., Kifayat K., Shi Q., Qureshi K.N., Askwith B. 2020. Efficient Non-Linear Covert Channel Detection in TCP Data Streams. IEEE Access. <https://doi.org/10.1109/ACCESS.2019.2961609>
- Preston R. 2019. DNS Tunneling Detection with Supervised Learning. 2019. IEEE International Symposium on Technologies for Homeland Security, HST 2019. <https://doi.org/10.1109/HST47167.2019.9032913>
- Salih A., Ma X., Petychev E. 2015. Detection and Classification of Covert Channels in IPv6 Using Enhanced Machine Learning. Proc. of The International Conference on Computer Technology and Information Systems (ICCTIS).
- Sherry J., Lan C., Popa R.A., Ratnasamy S. 2015. BlindBox: Deep packet inspection over encrypted traffic. SIGCOMM 2015 - Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication. <https://doi.org/10.1145/2785956.2787502>
- Shvartzshnaider Y., Wies T., Pavlinovic Z., Lakshminarayanan, Mittal P., Balashankar A., Nissenbaum H. 2019. Vaccine: Using contextual integrity for data leakage detection. The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019. <https://doi.org/10.1145/3308558.3313655>
- Steadman J., Scott-Hayward S. 2018. DNSxD: Detecting Data Exfiltration over DNS. 2018 IEEE Conference on Network Function Virtualization and Software Defined Networks, NFV-SDN 2018. <https://doi.org/10.1109/NFV-SDN.2018.8725640>
- Tsikerdekis M., Waldron S., Emanuelson A. 2021. Network anomaly detection using exponential random graph models and autoregressive moving average. IEEE Access. <https://doi.org/10.1109/ACCESS.2021.3116575>
- Vekshin D., Hynek K., Cejka T. 2020. DoH Insight: Detecting DNS over HTTPS by machine learning. ACM International Conference Proceeding Series. <https://doi.org/10.1145/3407023.3409192>
- Wendzel S., Link F., Eller D., Mazurczyk W. 2019. Detection of size modulation covert channels using countermeasure variation. Journal of Universal Computer Science. <https://doi.org/10.3217/jucs-025-11-1396>
- Yang P., Li Y., Zang Y. 2020. Detecting DNS covert channels using stacking model. China Communications. <https://doi.org/10.23919/JCC.2020.10.013>

- Yu B., Smith L., Threefoot M., Olumofin F. 2016. Behavior analysis based DNS tunneling detection and classification with big data technologies. IoTBD 2016 - Proceedings of the International Conference on Internet of Things and Big Data. <https://doi.org/10.5220/0005795002840290>
- Zhan M., Li Y., Yu G., Li B., Wang W. 2022. Detecting DNS over HTTPS based data exfiltration. Computer Networks. <https://doi.org/10.1016/j.comnet.2022.108919>

Den här rapporten presenterar en systematisk litteraturgenomgång om detektion av illasinnad nätverksbaserad dataexfiltrering. Sådan exfiltrering utgörs av angripares överföring av information från målmaskin till angriparmaskin. Eftersom det är mycket svårt att hålla alla angripare ute ur nätverken är detektionen av exfiltrering en viktig del i försvaret.

Litteraturgenomgången inkluderar 48 forskningsartiklar utgivna 2012–2022. Genomgången visar att artiklarna har tydliga syften men däremot oftast otydligt beskrivna hypoteser. Därtill visar genomgången att artiklarna fokuserar på labbmiljöer som efterliknar universitetsnät medan bara en artikel har en militär nätverksmiljö. Artiklarna beskriver i undantagsfall vilka data exfiltreringen rör. Det rör sig då om kreditkortsuppgifter, inloggningsuppgifter och dokument. Det är också ovanligt att artiklarna beskriver tänkta hotaktörer.

Artiklarna försöker detektera exfiltrering som döljs på olika sätt. Oftast rör det sig om döljande i form av placering av data i pakethuvuden, användning av protokoll som normalt inte nyttjas för användares dataöverföring eller av kryptering. Artiklarnas exfiltrering genomförs oftast via protokollet DNS, vilket har sitt legitima bruk i översättning av domännamn till IP-adresser.

Detektionsmetoderna baseras på olika förändringar som exfiltreringen ger upphov till. Det rör sig om nätverksmässiga skillnader i entropi, tidsaspekter, stränglängder, trafikflöden samt pakethuvudinnehåll. Algoritmerna är vanligen baserade på djupinlärning eller traditionell maskininlärning. Detektionsmetoderna utvärderas relativt knapphändigt i artiklarna.

Rapportens bedömning är att mer forskning behövs om framförallt detektion av exfiltrering som sker via videomöten, blockkedjenätverk, DNS över HTTPS, IPv6 och andra nyare protokoll. Därtill behövs forskning som snarare än att försöka detektera specifika tekniker tar fram mer generella algoritmer som kan detektera fler exfiltreringstekniker.