Aktivt skydd i cyberdomänen

En kunskapsöversikt

Henrik Karlzén, John Ziegenbein och Christian Vestlund

FOI-R--5797--SE

November 2025

FOI

Henrik Karlzén, John Ziegenbein och Christian Vestlund

Aktivt skydd i cyberdomänen

En kunskapsöversikt

Titel	Aktivt skydd i cyberdomänen – En kunskapsöversikt
Title	Active protection in the cyber domain – A literature review
Rapportnr/Report no	FOI-R--5797--SE
Månad/Month	November
Utgivningsår/Year	2025
Antal sidor/Pages	49
ISSN	1650-1942
Uppdragsgivare/Client	FMV
Forskningsområde	Cyberförsvar och cybersäkerhet
FoT-område	Inget FoT-område
Projektnr/Project no	E3800902
Godkänd av/Approved by	Emil Hjalmarson
Ansvarig avdelning	Cyberförsvar och ledningsteknik

Bild/Cover: TT Nyhetsbyrån, imageBroker/Norbert Acthelik

Sammanfattning

Aktivt cyberskydd utgörs av det urval av skyddstekniker för it-system som har en hög nivå av automatisering eller är helt autonoma. Denna studie utforskar de senaste fem årens utveckling inom aktivt cyberskydd i akademiska forskningsartiklar. Dessutom beskrivs tidigare FOI-forskning samt vissa cyberskydd som erbjuds av kommersiella företag. Studien visar att nya cyberskydd utvecklas inom både akademi och industri. Fokus i den akademiska litteraturen är främst på detektionstekniker medan kommersiella lösningar i större utsträckning också täcker in åtgärder som att isolera eller avlägsna hotaktörer. Den akademiska forskningen har generellt en låg mognadsgrad medan de kommersiella produkterna saknar oberoende praktiska utvärderingar. Det finns därmed frågetecken kring föreslagna och erbjudna skydds faktiska nytta.

En viktig aspekt är att en majoritet av de akademiska och kommersiella cyberskydden fokuserar på traditionella it-system. Sådana cyberskydd kan passa dåligt i fältnära militära miljöer. I dessa miljöer är det kritiskt att cyberskydden inte äventyrar den militära förmågan, vilken finns för att skydda människoliv och försvara landet. Ytterligare forskning behövs därför för att utvärdera vilka cyberskydd som passar i dessa sammanhang.

Nyckelord: cybersäkerhet, cyberförsvar, autonomi, aktivt cyberskydd

Summary

Active cyber protection constitutes the selection of IT system protection mechanisms that have a high level of automation, or are completely autonomous. This study explores the last five years of development within active cyber protection in academic research. Additionally, earlier FOI research is summarised, as are some cyber protection mechanisms provided by commercial companies. The study shows that new types of cyber protection mechanisms are developed both within academia and industry. The focus in the academic literature is mostly on detection techniques, while commercial solutions to a greater extent also cover response actions such as isolation and eviction of threat actors. The academic research generally shows a low level of maturity, while the commercial products lack independent practical evaluation. Consequently, there are uncertainties regarding the actual benefit of the proposed and offered protection mechanisms.

A vital aspect is that the majority of academic and commercial cyber protection mechanisms focus on traditional IT systems. These protection mechanisms may not be suitable in field-oriented military environments. In these environments it is critical that the cyber protection mechanisms do not compromise the military capability, which exists to save human lives and defend the country. More research is therefore needed to evaluate which cyber protection mechanisms that are suitable in these situations.

Keywords: cybersecurity, cyber defence, autonomy, active cyber protection

Innehållsförteckning

1 Inledning

Människor har sina begränsningar. Det är svårt för människor att vara på flera ställen samtidigt, att hålla stora mängder information i huvudet samt att hålla fokus i timmar för att sedan agera på sekunder. Människor saknar förutsättningar för att skydda utspridda it-system bestående av enorma datamängder som kan utsättas för angrepp vid oberäkneliga tillfällen.

Människor är dock uppfinningsrika. När den mänskliga kroppen inte räcker till tar människor fram verktyg för att lösa problemet. För det nyss nämnda problemet utgörs sådana verktyg främst av tekniska lösningar som delvis automatiskt, eller helt autonomt, kan skydda it-systemen. Sådan automatisering är särskilt lämplig för att både möta det växande hotet från angripare och för att nyttja de senaste årens framsteg inom AI.

Denna rapport beskriver tekniker som kan användas för att implementera aktivt cyberskydd i it-system. I rapporten används begreppet aktivt cyberskydd för att beskriva det urval av skyddstekniker för it-system som har en hög nivå av automatisering, eller rentav är autonoma. Studien undersöker teknikerna utifrån litteratur om teknikerna, men gör inga praktiska utvärderingar.

1.1 Syfte och mål

Rapporten syftar till att öka kunskapen om tekniker som kan användas för att implementera aktiva cyberskydd i Försvarsmaktens it-system. Den andra halvan av syftet kan brytas ned i fyra delar: 1) *implementera*, 2) *aktiva*, 3) *skydd*, 4) *i Försvarsmaktens it-system*. Nedan beskrivs mer om dessa fyra delar. För varje del beskrivs vilka aspekter som ingår i studien och vilka närliggande aspekter som inte ingår, utan istället avgränsas.

För att uppnå den åsyftade kunskapshöjningen är rapportens mål att presentera en sammanställning av relevanta tekniker, utifrån både akademisk och icke-akademisk litteratur.

1.1.1 Implementera

Vad gäller *implementera* så fokuserar studien på de tekniker och verktyg som existerar idag eller som kommer att vara relevanta inom en snar framtid. Detta innebär att fokus ligger på tekniker och verktyg som har en hög nivå av teknisk mognadsgrad (TRL, eng. technology readiness level). I vissa fall kan visserligen koncept (TRL 1–2) vara relevanta att studera om de förväntas ha bred påverkan framöver. Främst gäller dock att teknikerna behöver ha lämnat konceptstadiet och nått någon nivå av experimentell utvärdering (TRL 3–4). Forskningslitteratur bör kunna nå denna nivå. Ännu bättre är om teknikerna även har nått nivån där

tester sker i verkliga system och miljöer (TRL 5–7), eller rentav blivit produkter (TRL 8–9).

1.1.2 Aktiva i bemärkelsen någorlunda autonoma

Vad gäller *aktiva* så undersöker studien skydd som har en hög nivå av automatisering, eller som rentav är autonoma (självständiga från människor). I studien ingår även aspekter som rör möjligheten för en människa (försvarare) att förstå det autonoma beslutsfattandet och agerandet samt avbryta det vid behov. Däremot ingår inte någon juridisk eller etisk analys om sådant som ansvarsutkrävande för de autonomt tagna besluten.

1.1.3 Skydd för att upprätthålla säkerheten

Vad gäller *skydd* så undersöker studien skydd av egna system men täcker inte tekniker och verktyg som utför handlingar utanför dessa system. Därmed ingår inte tekniker som rör motåtgärder i andras system, såsom hackback och vita maskar. Inte heller ingår beaconing, vilket är en sorts larmfunktioner som följer med objekt till andras nätverk (Center for Cyber & Homeland Security, 2016). Begreppsligt gör rapporten ingen åtskillnad mellan skydd och de likartade termerna försvar, säkerhetsåtgärd, säkerhetsfunktion och säkerhetsmekanism. Mer om begrepp tas upp i bakgrundskapitlet.

1.1.4 Försvarsmaktens it-system till skillnad från annat

Vad gäller *i Försvarsmaktens it-system* så omfattar studien tekniker som är avsedda för militära it-miljöer; som har komponenter i fältnära miljöer snarare än i kontorsmiljöer samt tekniker som kan anpassas till sådana miljöer. Fältnära miljöer kan till exempel vara platser där ledningsfordon och handburna it-enheter används. Eftersom studien inte utgår från ett specifikt it-system inkluderas tekniker som bedöms kunna användas på fältnära system. Vad gäller angripare och angreppstyper görs inga avgränsningar utöver att det ska röra sig om angrepp som i huvudsak sker i cyberdomänen. Avgränsat är därmed fysiska (kinetiska) angrepp mot it-system. Ingen avgränsning görs i förhållande till konfliktspektrumet (fred, kris, krig).

1.2 Forskningsfrågor

Följande forskningsfrågor besvaras i rapporten:

- Vilka tekniker används för att implementera aktiva cyberskydd i it-system?
- Vilka för- och nackdelar har de identifierade teknikerna?
- Vilka förutsättningar finns för att kunna använda de identifierade teknikerna?

1.3 Läsanvisning

Den kunskap som rapporten syftar till att öka är relevant för flera olika kategorier av personal i framförallt Försvarsmakten och FMV men även till viss del hos civila aktörer. Det rör sig om personal som utvecklar cyberförmågor, planerar operationer eller använder it-system i operationer. Rapporten är tänkt att kunna läsas i sin helhet av dessa grupper.

Rapportens upplägg är som följer. Kapitel 2 ger en bakgrund med fokus på begrepp och då framförallt begreppet aktivt skydd samt beskriver tidigare FOI-rapporter. Kapitel 3 beskriver forskningsmetoden. Kapitel 4 och 5 beskriver forskningsresultaten från genomgången om den akademiska litteraturen respektive de kommersiella lösningarna. Kapitel 6 diskuterar resultaten och metoden. Kapitel 7 presenterar rapportens slutsatser. Kapitel 8 innehåller referenslistor.

2 Bakgrund

Det finns ingen entydig definition av begreppet *aktivt cyberskydd*. I denna rapport definieras det som det urval av skyddstekniker för it-system som har en hög nivå av automatisering, eller rentav är autonoma. I andra sammanhang varierar definitionen och det finns även närliggande begrepp. Det finns också mer konkreta begrepp som beskriver konkret implementation av olika aktiva skydd i till exempel intrångsdetektionssystem. Avsnitten 2.1 och 2.2 beskriver dessa olika begreppsvarianter djupare. Dessutom beskrivs i avsnitt 2.3 ett urval av tidigare FOI-forskning för att sätta den nuvarande rapporten i ett sammanhang och för att vägleda vidare läsning.

2.1 Begreppet aktivt skydd

Definitionerna av *aktivt skydd* skiljer sig åt i litteraturen, oavsett om det gäller cyberdomänen eller andra domäner. Avsikten i detta avsnitt är inte att avgöra exakt vad som betyder vad, utan snarare att ge läsaren en allmän inblick i terminologin på området och att samtidigt lyfta fram avsaknaden av en entydig definition av *aktivt skydd*. En tidigare FOI-rapport (Zouave, 2020) beskriver också en otydlighet i definitionerna inom det svenska cyberförsvaret och säger bland annat att källorna ”som utmejslar innebörden av ett aktivt cyberförsvar” bör läsas med försiktighet på grund av ”begreppsapparatens otydliga karaktär och dess applicering på en föränderlig miljö”.

Begreppet *aktivt skydd* kan delvis förstås genom att jämföra med närliggande begrepp. Exempelvis kan aktivt skydd jämföras med aktivt försvar; aktivt jämföras med passivt samt skydd jämföras med försvar och motåtgärder. Därtill finns det andra närliggande begrepp såsom proaktivt och reaktivt. Det finns också motsvarande begrepp på andra språk, som till exempel engelskans protection, defence, security och countermeasures.

I Försvarsmaktens (2024a) doktrinansats för cyberförsvar görs indelningen i *skydd* och *försvar*. *Skydd* beskrivs bland annat utgöras av ”åtgärder för att upprätthålla och stärka cybersäkerheten”. *Försvar* beskrivs bedrivas ”genom offensiva eller defensiva cyberoperationer” och ”ofta för att möta ett specifikt hot”. Detta kan jämföras med Svensk ordboks definitioner (Svenska Akademien. 2021a och 2021b) av *skydd* som ”anordning som förhindrar skadliga verkningar av oönskad företeelse” och av *försvar* som ”skydd (för någon eller något) genom motstånd mot angrepp”. I doktrinansatsen används inte begreppen *aktivt skydd* och *aktivt försvar*. Utifrån givna definitioner kan *försvar* dock ses som mer aktivt än skydd.

I ordlistan på Försvarsmaktens webbplats (Försvarsmakten, u.å.) förekommer begreppet *aktivt försvar* som en övergripande beskrivning av verksamheten: ”I dag måste Försvarsmakten kunna förstå, förutse, avhålla och möta – samtidigt. Det är det som kallas ett aktivt försvar. Det aktiva försvaret bygger på att Försvarsmakten konstant har en lägesuppdatering. Försvarsmakten analyserar vad som sker, är aktiva, proaktiva och vidtar åtgärder.” Försvarsmaktens webbplats beskriver också *aktivt försvar* som ”förmågan att verka aktivt, proaktivt och flexibelt” (Försvarsmakten, 2022).

Begreppet *aktivt skydd* förekommer också i en del av Försvarsmaktens dokument. När Försvarsmakten (2013) beskrev försvar mot ballistiska robotar och kryssningsrobotar delades försvaret in i två former av skydd: *aktivt* och *passivt*. Det aktiva skyddet beskrevs i termer av upptäckt med radar följt av vapenverkan med luftvärnsrobotar. Det passiva skyddet beskrevs i termer av ”fortifikatoriskt skydd, telekrigföring och spridning av skyddsvärda resurser”. Därtill nämndes att ”ett rörligt operativt och taktiskt uppträdande ökar skyddet.” På motsvarande sätt beskriver Försvarsmaktens koncept för defensiva cyberoperationer (DCO) (Försvarsmakten, 2024b) begreppet *passiv förmåga* som ”en enkelriktad koppling till terrängen som möjliggör observation men inte interaktion och påverkan av terrängen. I motsats till *aktiv förmåga*.” (källans kursivering).

I internationell litteratur på engelska finns motsvarande begrepp men ibland med lite annan innebörd. En relevant rapport (Center for Cyber & Homeland Security, 2016) om cyberhot och *active defense* togs fram av centret CCHS vid ett amerikanskt universitet, i samarbete med flera tidigare chefer vid departementet för inrikes säkerhet (DHS). Rapporten beskriver att begreppen aktivt och passivt försvar togs fram separat från varandra och då inom traditionella militära domäner. *Passivt försvar* kunde ge begränsat skydd mot en motståndare utan att kräva militär inblandning (eng. engagement). Exempelvis ingick bunkrar och andra skydd som gjorde att motståndare behövde mer resurser för att nå sitt mål. *Aktivt försvar* var däremot inriktat på mobilitet. Synsätten på aktivt och passivt försvar stämmer därmed ganska väl med den beskrivning från Försvarsmakten som gavs ovan. Därtill beskriver den amerikanska rapporten att översättningen till cyberdomänen varit svår eftersom det saknas perfekta motsvarigheter jämfört med de traditionella fysiska domänerna. Rapporten ger exempel från SANS Institute där *passivt försvar* handlar om att bygga arkitekturen så att systemet klarar sig mer på egen hand medan *aktivt försvar* fokuserar på analysarbete. Den amerikanska rapporten ser *aktivt försvar* som något proaktivt som samlar in information om motståndare, utfärdar sanktioner eller tekniskt interagerar med motståndare (men utan att bli offensivt). Som *passivt försvar* räknas grundläggande skydd som brandväggar och antivirus, medan *aktivt försvar* inkluderar sådant som vilseledning, hotjakt och informationsinsamling samt sanktioner, nedtagande av botnät och återtagande av tillgångar.

Den här rapporten fokuserar på *aktivt skydd*. Detta innebär att rapporten har mer fokus på analys (*aktivt*) än av arkitektur (*passivt*) samt mer fokus på egen terräng (*skydd*) än verkan utanför (*försvar*). I den amerikanska CCHS-rapportens terminologi exkluderas *passivt försvar* men också de mer aggressiva delarna av *aktivt försvar*. Resten av rapporten gör ingen åtskillnad mellan *skydd* och *försvar*.

2.2 Begrepp i ramverk och kommersiella produkter

Det finns flera sätt att kategorisera tekniker som faller inom begreppet skydd. I den här rapporten används Mitres D3fend-ramverk (Mitre, 2025) för att klassificera kommersiella produkter för aktivt skydd. D3fend har sju övergripande taktiker. En rör grundläggande modellering och tas inte upp mer i denna rapport. De andra taktikerna definieras enligt följande (fritt översatt):

- Härda – Ökar kostnaden för en motståndare att angripa it-system.
- Detektera – Identifierar fientlig åtkomst och obehörig aktivitet i it-system.
- Isolera – Skapar logiska eller fysiska barriärer i it-system för att minska möjligheterna för motståndare att ta sig in i andra delar av system.
- Vilseleda – Tillkännager, lockar och tillåter angripare åtkomst till en övervakad eller kontrollerad miljö.
- Avlägsna – Tar bort motståndare från ett nätverk.
- Återställa – Ställer tillbaka systemet till ett bättre tillstånd.

Varje taktik grupperar tekniker på flera nivåer. Exempelvis innehåller taktiken detektera sju övergripande tekniker, däribland filanalys, processanalys och nätverkstrafiksanalys.

Inom den kommersiella sektorn förekommer en stor mängd andra begrepp för olika tekniker. Begreppen är inte alltid tydligt definierade eller differentierade från närliggande begrepp. Därtill utvecklas produkter till att över tid innefatta mer funktionalitet eller förmågor. Vissa aktörer introducerar då nya termer för att beskriva den utvecklade produkten, medan andra aktörer behåller tidigare benämning. Detta leder till en begreppsförvirring som gör att det finns viss otydlighet i vad de olika termerna innefattar. Det finns inte utrymme att reda ut begreppsförvirringen i denna rapport. Mer begreppsförklaringar för de kommersiella produkterna ges dock i samband med att de beskrivs i rapportens resultatdel.

2.3 Tidigare FOI-rapporter

I detta avsnitt beskrivs ett urval av tidigare FOI-forskning inom området. Beskrivningarna ges både för att sätta den nuvarande rapporten i ett sammanhang och för att vägleda vidare läsning. Det som tas upp här är självläkande mjukvara, moving target defence (MTD), kontinuerlig beteendebaserad autentisering, lägesuppfattning, honungsfällor, detektion av exfiltrering samt intrångsdetektion och respons.

2.3.1 Självläkande mjukvara

Rodhe m.fl. (2014) ger en introduktion till självläkande mjukvara *”som en del av visionen om det autonoma datorsystemet”*. Mjukvaran är avsedd att på egen hand hantera nya typer av hot och fel. Sådan mjukvara *”är av särskilt intresse för verksamhetskritiska system och kan tillämpas i system som kräver hög nivå av autonomi, såsom rymdsonder och obemannade farkoster”*. De tillgängliga lösningarna fokuserade på detektion och diagnos av fel, medan den hanterande och återställande delen av lösningarna saknades.

2.3.2 Moving target defence

Holm m.fl. (2014) kartlade forskningsartiklar om moving target defence (MTD), det vill säga *”tekniker som försöker uppnå säkra system genom att ständigt förändra var sårbarheter befinner sig istället för att motarbeta deras existens”*. Rapporten beskriver att *”säkerheten för det skyddade systemet ges i och med att attackeraren inte hinner identifiera var sårbarheter för en konfiguration finns innan denna konfiguration byts ut”*. Själva existensen av MTD kan vara antingen känd eller okänd för angriparen. Exempel på MTD som ges är bland annat *”dynamisk förändring av IP-adresser i nätverk för att försvåra spaningsarbetet för en attackerare”* samt *”rörliga minnesadresser för mjukvaror för att försvåra injektion av skadlig kod i dem”*. Totalt identifierades sex typer av MTD: *”1) rörlig kodtransformering, 2) rörlig minnesallokering, 3) rörliga applikationer, 4) rörliga maskiner, 5) rörliga nätadresser och 6) kombinationer av dessa fem områden”*.

Därtill beskrivs att området funnits länge, även om begreppet var någorlunda nytt vid rapportens publicering. Det beskrivs att forskningsartiklarna på området hellre föreslår nya skydd än utvärderar dem: *”Fokus för de identifierade artiklarna är på introduktion och presentation av skydd snarare än utvärdering av deras prestanda eller hur de erbjuder ökad säkerhet. Sådana egenskaper diskuteras ofta med teori och/eller simuleringsresultat som stöd, men i de allra flesta fall utan empiri och under orealistiska förutsättningar. Detta gör att det är svårt att veta hur effektiva skydden faktiskt är och hur mycket de påverkar användarens upplevelse.”* Detta är ganska typiskt för forskningsartiklar och framförallt inom ett mindre moget forskningsområde. Därtill beskriver rapporten

hur forskningen kan göras mer mogen: *”Framtida arbete inom området bör fokusera mer på att utvärdera kvaliteten för förespråkade skydd, i synnerhet angående hur användaren påverkas och vilken hotmodell som är realistisk. Empiriska tester under realistiska förutsättningar, exempelvis med hjälp av så kallade cyber ranges, är ett sätt att uppnå mer valida resultat.”*

2.3.3 Kontinuerlig beteendebaserad autentisering

Karlzén m.fl. (2020) beskriver en introduktion till, och litteraturstudie av, en variant av biometrisk autentisering som baseras på användarnas beteende. Sådan autentisering kan exempelvis reagera på avvikelser från en användares normala tangentbordsanvändning, ögonrörelser, position, gångstil eller avvikelser som rör enhetens resursförbrukning. Jämfört med lösenord kan beteendebaserad biometrisk autentisering ske mer kontinuerligt och liknas vid intrångsdetektionssystem.

En av rapporternas slutsatser är att forskningslitteraturen framstår som ofullständig: *”Det finns många förslag på beteendebaserad biometrisk autentisering och kombinationer av sådana autentiseringsmetoder. Skillnader mellan olika användares sårbarhet har också lyfts fram och det till skillnad från övriga typer av autentisering. Men det saknas en rigorös teoretisk grund och seriösa utvärderingar med realistiska angripare, bedömningar av implementationskostnader, fullständiga metriker för utvärdering, representativa och stora grupper studiedeltagare och ordentlig jämförbarhet mellan studier. Andelen falska negativa är också generellt hög vilket påverkar de legitima användarnas möjlighet att lyckas autentiseras.”*

2.3.4 Lägesuppfattning

Bildsten m.fl. (2020) beskriver en förstudie om cyberlägesbild. Rapporten beskriver cyberlägesbild (eng. cyber situation(al) awareness) som *”en bild av aktuell (säkerhets)status inom cybermiljön. Att nå en lägesbild sker i tre steg: inhämtning av information, analys av information för att förstå nuläget samt förutsägelse av framtida händelser”*. Därtill beskrivs att åstadkommandet av en cyberlägesbild kräver att både kognitiva och tekniska utmaningar övervinns. *”Exempelvis måste verktyg för cyberlägesbilder samla in data från många olika delar av ett IT-system (som antiviruslarm och sårbarhetsskanningar) och presentera den insamlade informationen på ett för människan förståeligt sätt. Däremellan sker olika typer av analyser som exempelvis kan baseras på anomalidetektion och maskininlärning”*. Det beskrivs också att cyberlägesbilder är av nytta för att *”lära känna normalläget för det egna IT-systemet samt det normala användarbeteendet. Ett normalläge för ett IT-system kan innefatta hur nätverkstopologin ser ut samt hur information flödar i nätverken. Det kan också innefatta hur användare verkar i systemen, exempelvis inloggning från vilka*

geografiska platser och vid vilka tidpunkter. Utifrån normalläget kan sedan avvikelser och pågående angrepp identifieras."

Exempel på hur lägesbilden kan användas vid pågående angrepp rör bland annat att övervaka *"händelseloggar efter nya installationer eller schemaläggningar som inte är planerade"* och att analysera nätverksdata för ovanliga flöden, genom att exempelvis leta efter *"enheter som skickar stora mängder data till en extern mottagare som inte brukar motta data"*. Rapporten beskriver också olika typer av verktyg som underlättar arbetet med att upprätthålla lägesbilden. Exempelvis nämns SIEM som *"har förmåga att inhämta och analysera loggar (både nära realtid och historiskt), spara loggarna (exempelvis för regelefterlevnad) samt skicka larm och rapporter"*. Därtill beskrivs Security Orchestration Automation and Response (SOAR) som *"samlar in loggar och larmhändelser från olika loggproducenter, där larmhändelser från SIEM kan vara en av dem"*. SOAR beskrivs också ha *"fokus på automation och att få olika systemverktyg att fungera tillsammans (som i en orkester)."* SOAR har *"även fokus på att ta fram färdiga svar på incidenter med hjälp av så kallade planer (eng. playbooks)"*.

2.3.5 Honungsfällor

Karlzén (2021) beskriver honungsfällor (eng. honeypot), vilket är skydd som *"med hjälp av vilseledning lockar till sig angripare och får dem att stanna kvar"*. Rapporten beskriver också att honungsfällor ska efterlikna de riktiga systemen och *"används för att förskona ägarens IT-system från angrepp samtidigt som lärdomar för framtiden kan dras om angreppen"*. Honungsfällor av olika typ nämns, med möjliga indelningar beroende på *"placering, nätverksroll, syfte, interaktionsnivå, samverkan, dynamik och omdirigeringsförmåga"*. Därtill beskrivs följande om forskningens mognad: *"Forskningen om honungsfällor har lett till att många olika honungsfällor tagits fram genom åren. Vad gäller mognaden hos dagens honungsfällor så är den tillräcklig för att möta enkla breddangrepp. Mognadsgraden är generellt sett lägre vad gäller avancerade riktade angrepp."* Rapporten lägger dock till att honungsfällor även *"använts för att upptäcka tidigare okända sårbarheter (eng. zero-days)"*.

Att lura angripare och inhämta lärdomar är positivt för försvararen men det finns också risker med att använda honungsfällor. Rapporten nämner bland annat att förvirring kan uppstå när välvilliga utomstående upptäcker honungsfällan men utan att inse att den är en honungsfälla. Därtill nämns att angripare som förstår att det är en honungsfälla kan börja vilseleda försvararen så att denna drar felaktiga lärdomar, eller rentav använder honungsfällan som en språngbräda in i produktionsmiljön. Utöver honungsfällor nämner rapporten kortfattat liknande skydd som exempelvis tjärgropar (eng. tarpits), där angripare fördröjs.

Bildsten m.fl. (2021) beskriver ett försök att skapa ett dynamiskt nätverk av honungsfällor, där honungsfällorna anpassar sig efter angriparens attackmönster. I försöket studerades den svåra utmaningen med hur en *”sömlös överflyttning av angripare från produktionsmiljön till en honungsfälla kan genomföras utan att angriparen upptäcker att så har skett”*. Preliminära resultat visade att *”det grundläggande problemet med att kopiera en maskin under drift och att styra om angriparens nätverkstrafik är fullt möjligt att lösa”*. Överflyttningen tar dock tid vilket ger ett märkbart avbrott för angripare på 4–15 sekunder. Rapporten drar slutsatsen att beroende på *”vilken aktivitet angriparen utför under tiden för avbrottet, kan dock tiden vara för lång för att lyckas vilseleda angriparen”*.

2.3.6 Detektion av exfiltrering

Karlzén och Valassi (2022) gjorde en systematisk litteraturgenomgång av forskningsartiklar utgivna 2012–2022 i ämnet detektion av illasinnad nätverksbaserad dataexfiltrering. Rapporten beskriver att sådan exfiltrering utgörs av *“angripares överföring av information från målmaskin till angriparmaskin”*. Angripare döljer exfiltreringen på olika sätt: *”Oftast rör det sig [i artiklarna] 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.”* Angående detektionen av exfiltreringen baseras den på *”olika förändringar som exfiltreringen ger upphov till. Det rör sig om nätverksmässiga skillnader i entropi, tidsaspekter, stränglängder, trafikflöden samt pakethuvudinnehåll.”* Detektionsalgoritmerna är *”vanligen baserade på djupinlärning eller traditionell maskininlärning”*.

Angående detektionsmetodernas mognad beskriver rapporten att detektionsmetoderna utvärderas relativt knapphändigt i artiklarna och då i labbmiljöer som efterliknar universitetsnät medan bara en artikel har en militär nätverksmiljö. Därtill nämns att artiklarna bara i undantagsfall beskriver vilka data exfiltreringen rör och de tänkta hotaktörerna. 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.

2.3.7 Intrångsdetektion och respons

Karlzén och Sommestad (2023) (en forskningsartikel framtagen vid FOI) sammanställde forskning om automatiska incidenthanteringslösningar. Lösningarna kategoriseras efter deras indata och utdata, vilket utgörs av vilka intrångssignaler som finns respektive vilken respons de utför. Artiklarnas otydlighet gjorde kategoriseringen svår. En slutsats var dock att de vanligaste typerna av indata var tillgångsinventering, plattformsmonitorering och nätverkstrafikanalys. Vanligaste utdata var kommandon för nätverksisolering, det vill säga att konfigurera om brandväggarna. Forskningsartiklarnas lösningar jämfördes med kommersiella verktyg och de senare fokuserade mer på analys av enskilda filer.

3 Metod

Metodvalen redovisas uppdelat för akademisk litteratur respektive kommersiella produkter. I båda fallen finns indelningar i faserna sökning, inkludering och exkludering samt extrahering och syntes.

3.1 Akademisk litteratur

Den akademiska litteraturen utgjorde studiens huvuddel och de flesta metodval skedde inom ramen för denna litteratur. Mer detaljer om dessa val finns nedan.

3.1.1 Sökning

Sökningarna efter akademisk litteratur (forskningsartiklar) genomfördes i Scopus i maj 2025 med komplettering i augusti 2025. Söksträngen byggdes upp av termer utifrån rapportens syfte. Termerna konkretiserades genom tidigare kunskap hos rapportens författare, studier av ramverket Mitre D3fend samt läsning av tidigare FOI-rapporter. Därtill utfördes pilotsökningar i Scopus samt i Google Scholar. Några av källorna i pilotsökningen var en figur i en rapport av Center for Cyber & Homeland Security (2016) samt dokumentet Försvarsmaktens doktrinansats för cyberförsvar (Försvarsmakten, 2024a). Söktermerna återges i tabell 1.

Tabell 1 Söktermer, där OR användes för att sätta ihop raderna per kolumn, medan AND användes mellan varje kolumn.

Område	**Automatisering**	**Skydd**	**Militär domän**
Cyber*	Auto*	Defense	Military
Network	Active	Defence	Cyber operation
		Incident response	IoBT
		Incident handling	Internet of battlefield things
		Countermeasure	

Söktermerna bildade en söksträng för att hitta forskningsartiklar som matchade söktermerna i olika sökfält. Söktermerna i kolumnen militär domän användes i alla Scopus sökfält, medan de övriga söktermerna användes i sökfälten titel, sammanfattning och nyckelord. Söksträngen gav 738 träffar i Scopus 19 augusti 2025. Den slutliga söksträngen såg ut som följer:

```
TITLE-ABS-KEY ( ( cyber* OR network ) AND ( auto* OR ACTIVE ) AND ( "defense"
OR "defence" OR "incident response" OR "incident handling" OR countermeasure ) )
AND ( military OR "cyber operation" OR iobt OR "internet of battlefield things" )
```

3.1.2 Inkludering och exkludering

Utifrån sökträffarna gjordes både automatisk och manuell filtrering.

För den automatiska filtreringen gjordes valet att enbart inkludera forskningsartiklar som klassificerats som datavetenskap (eng. computer science) och som skrivits på engelska. Därtill inkluderades enbart artiklar publicerade tidigast 2019. Dessutom exkluderades de artiklar som i databasen klassificerats som brev (eng. letter), konferenssammanställning (eng. conference review; inte att förväxla med litteraturgenomgång) eller som dragits tillbaka (eng. retracted).

För den manuella filtreringen av sökträffarna användes en uppsättning inkluderingskriterier. Dessa kriterier togs fram gemensamt bland rapportförfattarna och utgick från rapportens syfte. För att en artikel skulle inkluderas behövde den beskriva en teknik som:

- Skyddar mot angrepp, men inte utgör motangrepp eller offensiva åtgärder.
- Är avsedd för militära it-miljöer eller miljöer som har komponenter i fältnära miljöer (snarare än i kontorsmiljöer) eller som kunde anpassas till sådana miljöer.
- Ha en högre mognadsgrad (TRL-nivå) än två på en niogradig skala, där högre nivå innebär mognare teknik.

Dessa kriterier användes för att bedöma om en artikel skulle inkluderas. Bedömning gjordes först på titel. För kvarvarande artiklar gjordes bedömning även på sammanfattning och därefter i förekommande fall även på fulltext. För varje bedömningssteg skapades en samsyn av kriterierna genom att en mindre delmängd av artiklarna slumpades fram och delades upp bland de tre forskarna som författat denna rapport, så att varje forskare bedömde två tredjedelar av de slumpade artiklarnas metadata eller fulltext. Avvikelser i bedömningarna diskuterades sedan för att fastställa kriterierna och förståelsen av dem. För varje bedömningssteg delades sedan resten av artiklarna upp mellan forskarna. Två (av tre) forskare läste varje artikels titel. En artikel inkluderades om minst en av forskarna ansåg den relevant utifrån en helhetsbedömning av kriterierna. Därefter lästes kvarvarande artiklars sammanfattningar. Även här bedömdes varje artikel av två forskare. För läsning av sammanfattningarna gjordes en uttrycklig bedömning av varje inkluderingskriteriums uppfyllnad. En artikel inkluderades om minst en av forskarna ansåg den relevant. De inkluderade artiklarna fulltextgranskades sedan, utom fem artiklar som forskargruppen saknade tillgång till (varför dessa artiklar istället ströks). Fulltextgranskning gjordes av 115 artiklar, uppdelat på en tredjedel per forskare.

Totalt bedömdes 55 artiklar vara relevanta att inkludera för extrahering och syntes. Den vanligaste anledningen till exkludering var att artikeln saknade fokus på cyberskydd. Den andra stora anledningen till exkludering var att tekniken som presenterades bedömdes ha en alltför låg mognadsgrad (TRL 1–2). Angående

kriteriet om fältnära miljöer inkluderades 32 artiklar (av de 55) där miljöerna visserligen inte var fältnära men där skyddsteknikerna bedömdes kunna anpassas till sådana miljöer.

3.1.3 Extrahering och syntes

Från den identifierade och bedömda litteraturen extraherades relevanta delar, utöver vad som redan extraherats i samband med bedömningarna. Extraheringen utgick i huvudsak från vad som uttryckligen nämndes i artiklarna, snarare än att göra en egen analys av artiklarna. Extrahering gjordes av följande aspekter, baserade på studiens forskningsfrågor:

- Typ av teknik.
- För- och nackdelar med tekniken, inklusive vilka hot som motverkas.
- Förutsättningar i form av typen av miljö.

Extrahering gjordes också för att skapa en uppfattning av artiklarnas kvalitet (exempelvis huruvida artikelns forskningsfråga är tydlig).

Först gjordes en kontroll av att extraheringsformuläret uppfattades lika bland forskarna som författat denna rapport. Därefter delades artiklarna upp bland forskarna så att extrahering av en viss artikel bara gjordes av en forskare.

Det extraherade materialet sammanställdes med enklare summeringar. Eftersom artiklarna skiljer sig ganska mycket åt gjordes inga större försök att skapa mer omfattande synteser. Det fanns till exempel ingen vits med att presentera medelvärden av alla artiklars detektionsförmåga. Syftet med litteraturgenomgången var dessutom primärt att ge en översikt av lämpliga tekniker än att säga vilken effekt teknikerna har tillsammans. Det relativt spretiga materialet gjorde också att det var svårt att hitta lämpliga taxonomier att kategorisera artiklarna med, varför en mer induktiv ansats användes i flera fall.

3.2 Kommersiella produkter

Litteratur om de kommersiella produkterna kompletterade den akademiska litteraturen. Mer detaljer om identifieringen av produkterna beskrivs nedan.

3.2.1 Sökning

För att identifiera relevanta kommersiella produkter gjordes sökningar på Google kompletterat med Gartners sammanställningar av produkter i vad de kallar Magic Quadrant (Gartner, u.å.). Enligt Pollock och Williams (2009) är Gartner en tongivande och inflytelserik aktör inom branschanalys, men inte nödvändigtvis oberoende från de tillverkare som bedöms. De produkter som redan identifierats vid sökningar på Google visade sig vara med i minst en Magic Quadrant. Produkter identifierades också utifrån företag som var kända av den här

rapportens författare sedan tidigare. Sådana företag var, med produkttyp i parentes, Clavister (NDR), L3 Harris (AMTD), Booz Allen Hamilton (SIEM) och BAE systems (CPS PP och XDR).[1]

3.2.2 Inkludering och exkludering

Gartners lista över Magic Quadrants (Gartner, u.å) granskades och de Magic Quadrant som inkluderades beskrivs i tabell 2. Magic Quadrants inkluderas om tekniken bedömdes innefatta säkerhet samt kunna appliceras i ett fältnära system. Molnbaserade tekniker bedömdes inte vara applicerbara och exkluderades därmed. Det är oklart hur Gartner skapar sina Magic Quadrants, vad gäller hur de samlar mätdata eller fattar beslut om vilka produkter som ska vara med utifrån mätdata.

Därtill inkluderades sex produkter från de företag som var kända av rapportförfattarna sedan tidigare. Listan av produkter gallrades utifrån samma inkluderingskriterier som för den akademiska litteraturen. Totalt bedömdes 83 kommersiella lösningar vara relevanta.

Tabell 2 Kolumnen ”Inkl.” avser antal inkluderade produkter av de som ingår i en Gartner Magic Quadrant. Kolumnen ”Exkl.” avser anledningen till exkludering.

Magic Quadrant	**Inkl.**	**Exkl.**	**Referens**
Endpoint Protection Platforms (även kallat Endpoint Detection and Response, EDR)	16/16	–	(Gartner, 2024a)
Security Information and Event Management (SIEM)	25/34	Molnbaserad	(Gartner, 2024b)
Observability Platforms (OP)	7/19	Managerad tjänst	(Gartner, 2024c)
Privileged Access Management (PAM)	6/9	Otillräcklig autonomi	(Gartner, 2024d)
CPS Protection Platforms (CPS PP)	13/17	Otillräcklig autonomi	(Gartner, 2025a)
Network Detection and Response (NDR)	10/11	Otillräcklig autonomi	(Gartner, 2025b)
Summa	77	–	–

[1] AMTD är automatiserad MTD (moving target defence). XDR är en utökning av EDR och NDR. CPS är cyberfysiska system medan PP står för protection platform (skyddsplattform).

En del produkter exkluderades då de inte bedömdes inneha tillräcklig autonomi för att betraktas som aktivt cyberskydd. Flera lösningar bedömdes inte vara relevanta för studien då de inte passade som fältnära lösningar eftersom de enbart erbjöd managerade tjänster snarare än fristående produkter. De lösningar som enbart erbjöd produkter för molntjänster exkluderades också. Dessa exkluderingar var huvudsakligen förekommande i kategorierna för SIEM samt OP.

3.2.3 Extrahering och syntes

Det är både enklare och svårare att utvärdera kommersiella lösningar än forskningsmässiga motsvarigheter. Det är enklare eftersom mognaden generellt sett är högre. Det är svårare eftersom en kommersiell aktör har mer intresse av att sälja en produkt än att förklara hur den fungerar. Av det senare skälet gjordes ingen utförlig extrahering. Den extrahering som gjordes baserades på samma källor som användes vid identifiering av produkten med tillägget produktens webbplats. Extraheringen rörde typen av teknik och avsedd miljö. Däremot extraherades inte information om för- och nackdelar med tekniken. Det extraherade materialet sammanställdes med enklare summeringar.

4 Akademisk litteratur

Detta kapitel redovisar resultaten för genomgången av den akademiska litteraturen. Referenser till de artiklar som inkluderades i litteraturstudien finns i avsnitt 8.2 i rapportens referenslista. I löptexten anges dessa artikel-referenser med ID-nummer.

4.1 Vilka tekniker används för att implementera aktiva skydd?

Tabell 3 ger en översikt över typerna av tekniker och hur många artiklar som använder tekniken. Artiklarnas tekniker benämns i huvudsak utifrån hur artiklarna själva benämner tekniken. Eftersom begreppen skiljer sig åt mellan artiklarna har den här rapportens författare till viss del ensat begreppen. Fokus har varit på att använda termer som är vedertagna inom cybersäkerhetsområdet, men inte nödvändigtvis som del av en etablerad gemensam taxonomi.

Tabell 3 Översikt över typerna av tekniker och antal artiklar som använder tekniken. Kolumnen med Antal detektion anger hur många av artiklarna som innehöll en funktion eller komponent för detektion.

Typ av teknik	**Antal artiklar**	**Antal detektion**
Nätverksbaserat intrångsdetektionssystem (NIDS)	13	13
Intrångsdetektionssystem (IDS)	7	7
Intrångspreventionssystem (IPS)	7	7
Värdbaserat intrångsdetektionssystem (HIDS)	3	3
Honungsfälla	7	2
Autonoma agenter	6	6
Ramverk för autonomt cyberförsvar	5	5
Moving target defence (MTD)	5	2
Spelteoretisk riskanalys	1	0
Lägesuppfattning för förutsägelser	1	0
Totalt	*55*	*45*

De allra flesta tekniker fokuserar på detektion i ganska allmänna termer, vilket löst kan benämnas intrångsdetektionssystem (IDS). IDS-lösningarna kan delas upp i huruvida de sker hos en enskild värddator (eng. host) vilket benämns HIDS, eller på nätverksnivå vilket benämns NIDS. Denna indelning gjordes när varianten tydligt framgick i artikeln och räknas då enbart som den specifika typen. Ibland har IDS-lösningarna också en reaktionsdel vilket ger upphov till ett intrångspreventionssystem (IPS). Som reaktionsdel räknas inte bara när tekniken

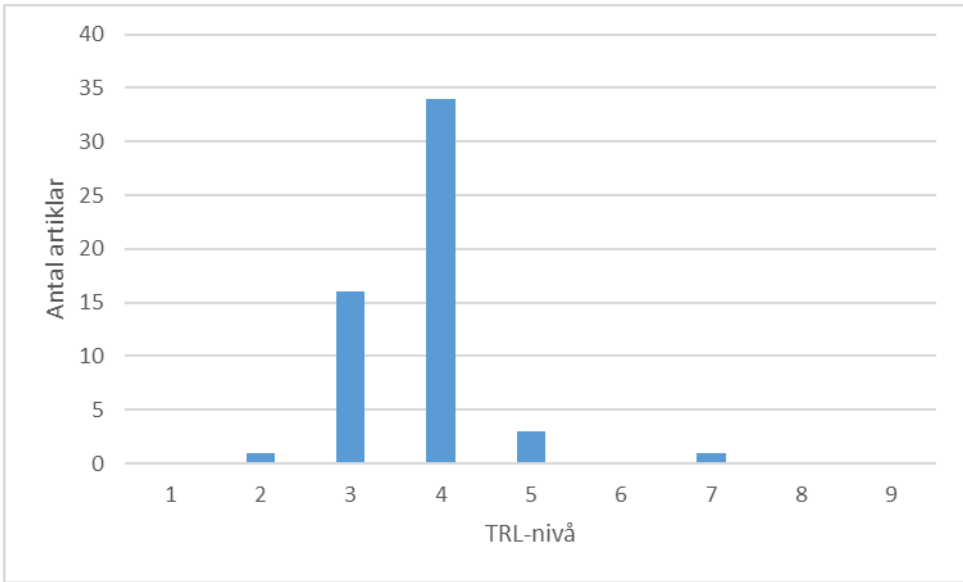
utförde åtgärder utan också när tekniken istället föreslog eller rekommenderade åtgärder utan att utföra dem. Alla HIDS inkluderade en IPS-komponent.

Förutom IDS-varianter och IPS förekommer även andra tekniker. Även dessa andra tekniker kan innehålla en funktion eller komponent för detektion. När så är fallet anges detta i den sista kolumnen i tabellen. Bland dessa andra tekniker finns de som baseras på moving target defence (MTD) samt honungsfällor. Det finns också autonoma agenter, vilka kan integrera information från flera typer av loggkällor för att detektera avvikande händelser. Ett exempel på en autonom agent beskrivs i artikeln med ID 555 (i referenslistan) och testades vid en Nato-övning.

Därtill finns ramverk för autonomt cyberförsvar. Ramverken rör arkitekturer för detektion, beslutsfattande och respons. Arkitekturerna använder sig av skyddstekniker som IDS, IPS, MTD och honungsfällor för att realisera skyddet på en systemnivå. Realiseringen av ramverken kan också ske genom användandet av autonoma agenter. Ett exempel på ramverk är Natos ramverk för AICA (Autonomous Intelligent Cyber Defense Agent), vilket omnämns i artikel 599.

Dessutom finns det en artikel med en teknik för spelteoretisk riskanalys samt en artikel med en teknik för lägesuppfattning som förutsåg framtida skeende, snarare än detektion i efterhand.

Att en artikel behandlar en teknik innebär inte att tekniken är redo att implementeras. Hur långt från implementering som tekniken befinner sig beror på mognadsgraden. Artiklarnas mognadsgrad bedömdes utifrån TRL, vilket är en ofta använd skala för hur mogen teknik är. Som kan skönjas av figur 1 bedömdes majoriteten (34 av 55) av artiklarna ha en mognadsgrad på TRL 4, vilket ungefär innebär att tekniken utvärderats i labbmiljö. Endast fyra artiklar bedömdes ha en högre nivå och det gäller tre artiklar med TRL 5 (med ID 233, ID 367 och ID 389, i referenslistan) och en artikel med TRL 7 (ID 7). Den senare artikeln utvärderade en teknik med honungsfällor i en operativ molnbaserad miljö. För artiklarna med lägre TRL än 4 fanns sexton artiklar med TRL 3 och en artikel med TRL 2 (ID 4). Den senare artikeln utvecklade ett koncept för spelteoretisk riskanalys och inkluderades som undantag från mognadskravet bland inkluderingskriterierna eftersom den hade tydlig militär fältnära miljö.



Figur 1 Artiklar per mognadsgrad (TRL).

4.2 Vilka för- och nackdelar har teknikerna?

Nedan beskrivs de hot som skyddsteknikerna är avsedda att möta. Senare i avsnittet beskrivs andra för- och nackdelar med teknikerna.

Tabell 4 kategoriserar de hot som artiklarnas tekniker skyddar mot. Det är vanligt att artiklarna berör mer än en typ av hot. Detta gäller ungefär hälften av artiklarna. Femton artiklar möter 5–9 hot (enligt denna rapports kategorisering), femton artiklar möter 2–4 hot, medan sexton artiklar möter 1 hot. För nio artiklar är hotet oklart eller nämns inte.

Tabell 4 Hoten som artiklarnas skyddstekniker möter, enligt artiklarna.

Typ av hot	**Antal**
Tillgänglighetsangrepp (DoS)	27
Rekognoscering och skanning	24
Skadlig kod	17
Exploatering	17
Fuzzing och kodinjektion	13
Utnyttjande av bakdörrar	11
Privilegiehöjning	7
Avlyssning	4
Botnät	3
Brute force	3
Spoofning	3

Typ av hot	Antal
Utpressningsprogram	2
Insider	1
Fysiskt angrepp för åtkomst	1
Fjärrunderhåll	1
Manipulation	1
Maskering	1
Drive-by-download	1
Nätfiske	1
Stulna inloggningsuppgifter	1
Man-in-the-middle	1
Spelteori där angripare anpassar sig till försvararens val av skydd	1

Beskrivningarna av hoten är ofta otydliga i artiklarna. Exempelvis nämns ofta hotrelaterade ord i förbifarten och utan förklaringar. I en del andra fall kommer angreppsdata från dataset som genererats av andra forskare. Men dataseten används då utan att ge mer än knapphändiga förklaringar av hoten (angreppen) i dataseten. Exempelvis har dataseten grova kategorier enligt en godtycklig, och rentav svårförståelig, terminologi. En artikel som utgör ett undantag med mer information om hot är artikeln med ID 88 (i referenslistan).

I artiklarna varierar det också ifall hoten beskrivs utifrån hur angrepp utförs eller vad deras syften är. Vissa av hoten säger också något om konsekvenserna för försvararen: tillgänglighetsangrepp (eng. denial of service) och utpressningsprogram påverkar tillgängligheten, medan avlyssning påverkar konfidentialiteten. Därtill beskriver artikeln med ID 53 hot mot tillgänglighet, riktighet, konfidentialitet, oavvislighet med mera. Dessutom kan skydd påverkas av hoten på olika sätt. Exempelvis påverkas (kringgås) skyddet behörighetskontroll av angreppstekniken privilegiehöjning.

Förutom hot tas andra för- eller nackdelar med teknikerna upp någorlunda tydligt av 36 artiklar. Dessa andra för- och nackdelar (aspekter) beskrivs i tabell 5. I vissa fall kan en aspekt vara både en fördel och en nackdel, vilket framgår av tabellens första kolumn. Aspekterna existerar i artiklarnas versioner av teknikerna. Det är möjligt att förädling av teknikerna kan komma till rätta med nackdelarna eller bygga vidare på fördelarna. Det är också möjligt att en riktig implementation i verklig miljö inte skulle klara av att ge samma fina siffror som uppnås i en labbmiljö.

Tabell 5. Sammanställda för- och nackdelar (aspekter) med teknikerna som beskrivits i forskningsartiklarna samt i hur många artiklar respektive aspekt förekommit. Plus och minus anger att raden rör fördelar eller nackdelar.

+/-	Aspekt	Antal	Exempel
+ -	Noggrannhet, precision samt andra typiska AI-metriker	14	Förhållandet mellan förmåga att detektera angrepp (sanna positiva) och att inte larma i onödan (falska positiva).
-	Begränsade angrepp	6	Svårigheten att upptäcka angrepp som liknar användarbeteende eller som utnyttjar okända sårbarheter.
-	Begränsad miljö	5	Orealistiskt simulerade användare i testfasen, vilket gör att mycket av det som riktiga användare gör kommer ses som avvikelser.
+	Komplett miljö (mer realistisk)	2	Förmåga att analysera krypterad trafik. Kompatibilitet.
-	Resurskrävande	11	Hög komplexitet på grund av mängden loggar som rör detektion och beslut. Det är svårt att välja respons samt förklara beslut.
+	Resurssnålt	4	Snabbhet samt skalbarhet. Att klara stora dataflöden och högdimensionella data.
-	Sänkt nätverksprestanda	4	Paketförluster.
+	Robusthet	3	Motståndskraft vid mer komplexa och dynamiska hot samt låg risk för missbruk (trots införande av honungsfällor).
+ -	Öppen källkodsmjukvara	1	Använder tillgänglig öppen källkodsmjukvara.

Två artiklar jämför dessutom olika AI-algoritmer (som komponenter i skyddstekniker). Den ena artikeln anger olika för- och nackdelar för de olika algoritmerna. Den andra artikeln beskriver mer allmänt att kombinationen av algoritmer kan ge bättre detektionsförmåga.

4.3 Vilka förutsättningar finns för att använda de identifierade teknikerna?

Den viktigaste förutsättningen för att kunna använda en skyddsteknik är att tekniken fungerar för den miljö (eller den typ av system) som ska försvaras. Vilka typer av miljöer som artiklarnas tekniker tagits fram för anges därför i tabell 6. I vissa artiklar angavs flera miljöer och i fyra artiklar framgick inte

miljön tydligt nog för att möjliggöra kategorisering. Av miljöerna klassificerades 23 av 55 som särskilt fältnära eller med militär koppling. Det rörde sig om de som var IoT eller cyberfysiska system (CFS) eller (i tre fall) som hade uttrycklig militär koppling.

Tabell 6 Miljötyperna som artiklarnas tekniker tagits fram för.

Miljötyp	**Antal**
Kontorsnät (enterprise)	30
Sakernas internet (IoT)	11
Cyberfysiska system (CFS)	6
Mjukvarudefinierade nätverk (SDN)	6
Moln	2
Wifi	1
Mobila ad-hoc-nätverk	1

Artiklarna om CFS väljer den här rapportens författare att lyfta fram särskilt. I dessa artiklar var fokus på: drönares kommunikation (för artikeln med ID 53 i referenslistan), ett fordonssystem (ID 56),[2] maritima system (ID 233), militärt trafikledningssystem (ID 4), system för energiförsörjning (ID 367) samt cyberfysiska system i allmänhet (ID 377).

Här lyfts också artiklar med militär koppling särskilt fram. En artikel med SDN-miljö hade militär koppling och det rörde sig om ett taktiskt koalitionsnät (ID 88). En artikel med kontorsnätsmiljö byggde även ett ad-hoc-nät för taktisk militär användning (ID 124). En annan artikel utvärderade i en Nato-övning en implementation av en agent för Natos referensarkitektur AICA (ID 555). Vid övningen användes kontorsnätsmiljöer. De autonoma agenterna är dock också tänkta att kunna användas i andra militära miljöer såsom obemannade drönare och marina industriella kontrollsystem (ID 599).

Förutom anpassning till rätt miljötyp finns det också andra förutsättningar för att kunna använda teknikerna. Det gäller framförallt mer konkret möjlighet att integrera tekniken i ett visst system och att samverka med andra tekniker. Dessa aspekter med integration och interoperabilitet beskrivs dock sällan i artiklarna. Ett undantag (ID 352) beskriver kortfattat att tekniken inte kräver modifiering av nätverket, exempelvis vad gäller mekanismer för intra-domän-routing och tunnling.

[2] Denna artikel skrevs av forskare med militär anknytning.

5 Kommersiella lösningar

Detta kapitel beskriver kortfattat de kommersiella lösningarna och hur lösningarnas tekniker kan kategoriseras.

5.1 Produkter och deras miljöer

Tabell 7 ger en översikt över kommersiella lösningar för aktivt cyberskydd. Mer detaljer om kategorierna av tekniker beskrivs i avsnitt 5.2.

Tabell 7 Kommersiella lösningar för aktivt cyberskydd.

Teknik	**Miljö**	**Företag**	**Produktnamn**
AMTD	IoT, Kontorsnät (enterprise)	L3 Harris	Agile guardian
CPS PP	IoT, CFS	Armis	Armis Centrix
CPS PP	IoT, CFS	BAE systems	Cyber-A2
CPS PP	IoT, CFS	Cisco	Cyber Vision
CPS PP	IoT, CFS	Claroty	Claroty CTD
CPS PP	IoT, CFS	Darktrace	OT
CPS PP	IoT, CFS	Forescout Technologies	OT Security
CPS PP	IoT, CFS	Fortinet	FortiGate Next Generation Firewall
CPS PP	IoT, CFS	Microsoft	Defender for IoT
CPS PP	IoT, CFS	Nozomi Networks	Guardian
CPS PP	IoT, CFS	OPSWAT	MetaDefender OT Security
CPS PP	IoT, CFS	Palo Alto Networks	Strata Network Security Platform
CPS PP	IoT, CFS	Radiflow	CPS Security platform
CPS PP	IoT, CFS	Tenable	Tenable OT Security
CPS PP	IoT, CFS	TXOne Networks	Edgeone
EDR	Kontorsnät (enterprise)	Bitdefender	GravityZone
EDR	Kontorsnät (enterprise)	Broadcom	Carbon Black Endpoint Security
EDR	Kontorsnät (enterprise)	Check Point Software Technologies	Check Point Harmony Endpoint
EDR	Kontorsnät (enterprise)	Cisco	Secure endpoint/XDR
EDR	Kontorsnät (enterprise)	Crowdstrike	Falcon

Teknik	Miljö	Företag	Produktnamn
EDR	Kontorsnät (enterprise)	Cybereason	Cybereason Defense Platform
EDR	Kontorsnät (enterprise)	ESET	ESET PROTECT
EDR	Kontorsnät (enterprise)	Fortinet	FortiClient/FortiXDR
EDR	Kontorsnät (enterprise)	Microsoft	Defender
EDR	Kontorsnät (enterprise)	Palo Alto Networks	Cortex
EDR	Kontorsnät (enterprise)	SentinelOne	Singularity for Endpoint
EDR	Kontorsnät (enterprise)	Sophos	EDR/XDR
EDR	Kontorsnät (enterprise)	Trellix	Endpoint Security
EDR	Kontorsnät (enterprise)	TrendMicro	Trend Vision One Endpoint Security
EDR	Kontorsnät (enterprise)	WithSecure	Elements Endpoint Security
NDR	Kontorsnät (enterprise)	Arista Networks	Arista NDR
NDR	IoT, CFS, Kontorsnät (enterprise)	Clavister	PASAD
NDR	Kontorsnät (enterprise)	Corelight	Open NDR Platform
NDR	Kontorsnät (enterprise)	Darktrace	NETWORK
NDR	Kontorsnät (enterprise)	ExtraHop	RevealX
NDR	Kontorsnät (enterprise)	Gatewatcher	Gatewatcher NDR Platform
NDR	Kontorsnät (enterprise)	Stellar Cyber	Stellar Cyber NDR
NDR	Kontorsnät (enterprise)	ThreatBook	TDP
NDR	Kontorsnät (enterprise)	Trellix	Trellix NDR
NDR	Kontorsnät (enterprise)	TrendMicro	Trend Vision One – Network Security

Teknik	Miljö	Företag	Produktnamn
NDR	Kontorsnät (enterprise)	Vectra AI	Vectra AI Platform
OP	Kontorsnät (enterprise)	BMC	BMC Helix Operations Management
OP	Kontorsnät (enterprise)	Dynatrace	Dynatrace platform
OP	Kontorsnät (enterprise)	Elastic	Elasticsearch, Logstash Kibana, Beats
OP	Kontorsnät (enterprise), Moln	IBM	Instana
PAM	Kontorsnät (enterprise)	ARCON	ARCON PAM
PAM	Kontorsnät (enterprise)	BeyondTrust	PASM with Pathfinder
PAM	Kontorsnät (enterprise)	Broadcom (Symantec)	Symantec PAM
PAM	Kontorsnät (enterprise)	CyberArk	CyberArk PAM
PAM	Kontorsnät (enterprise)	ManageEngine	PAM360
PAM	Kontorsnät (enterprise)	Netwrix	Netwrix PAM
SIEM + SOAR	Kontorsnät (enterprise)	Booz Allen Hamilton	Darklabs Detect Darklabs Protect
SIEM + SOAR	Kontorsnät (enterprise)	Devo	Security Data Platform
SIEM + SOAR	Kontorsnät (enterprise)	Elastic	SIEM from Elastic Elastic SOAR
SIEM + SOAR	Kontorsnät (enterprise)	Exabeam	LogRhythm SIEM Exabeam SOAR
SIEM + SOAR	Kontorsnät (enterprise)	Fortinet	FortiSIEM FortiSOAR
SIEM + SOAR	Kontorsnät (enterprise)	Gurucul	Gurucul Next gen SIEM Gurucul Reveal
SIEM + SOAR	Kontorsnät (enterprise)	IBM	IBM Qradar SIEM IBM Qradar SOAR
SIEM + SOAR	Kontorsnät (enterprise)	Logpoint	Logpoint SIEM Logpoint SOAR
SIEM + SOAR	Kontorsnät (enterprise)	ManageEngine	Log360

Teknik	Miljö	Företag	Produktnamn
SIEM + SOAR	Kontorsnät (enterprise)	Microsoft	Microsoft Sentinel
SIEM + SOAR	Kontorsnät (enterprise)	Odyssey	ClearSkies
SIEM + SOAR	Kontorsnät (enterprise)	QAX	QAX SIEM QAX SOAR
SIEM + SOAR	Kontorsnät (enterprise)	Rapid7	InsightIDR InsightConnect
SIEM + SOAR	Kontorsnät (enterprise)	Securonix	Securonix Unified Defense SIEM Securonix SOAR
SIEM + SOAR	Kontorsnät (enterprise)	Splunk	Splunk Enterprise Security Splunk SOAR
SIEM + SOAR	Kontorsnät (enterprise)	Venustech	Venusense Unified Security Management
XDR	IoT, CFS	BAE systems	Resilience in Depth

Produkterna är i huvudsak utvunna ur Gartners sammanställningar benämnda Magic Quadrant. Vad gäller de produkter som inkluderas här finns det flera företag med produkter i flera Magic Quadrant. De vanligaste företagen är Microsoft och Fortinet (tre produkter vardera). Det finns även tretton företag som har produkter i två Magic Quadrant. För dessa siffror räknas inte SIEM och SOAR som separata produkter.

5.2 Olika typer av skydd

Termerna som de kommersiella aktörerna använder kan kopplas till Mitres D3fend-ramverk enligt tabell 8. Denna uppdelning är rapportförfattarnas tolkning av termerna och syftar till att gå från de ganska svårförståeliga kommersiella begreppen (i första kolumnen) till mer lättförståeliga försvarstaktiker (övriga kolumner).

Tabell 8 Termer för typer av kommersiella lösningar och vad de täcker. Grön cell med ifylld ring ● betyder primärt syfte; gul cell med delvis streckad ring ○ betyder sekundärt syfte; röd cell med helstreckad ring ○ betyder stöds ej.

Term	Härda	Detektera	Isolera	Vilseled	Avlägsna	Återställ
SIEM	○	●	○	○	○	○
SOAR	○	○	●	○	●	●
OP	●	●	●	○	○	○
PAM	●	●	●	○	●	○
NDR	●	●	●	○	●	○
EDR	●	●	●	○	●	○
XDR	●	●	●	○	●	●
CPS PP	●	●	●	○	●	●
AMTD	●	○	○	●	○	●

SIEM har autonomi i form av intrångsdetektion samt användar- och enhetsbeteendeanalys. Därtill ges autonomt stöd till säkerhetscenter för att underlätta prioritering och informationsförädling för identifierade sårbarheter. Eftersom det sannolikt är svårare att få tag på expertis i fält finns det ett särskilt värde i att automatisera och effektivisera detta arbete för fältnära system. För samtliga SIEM-produkter erbjuder tillverkarna också en SOAR för autonom respons.

Observerbarhetsplattformar (OP) beskrivs av Gartner som produkter som konverterar inhämtade mätdata till insikter och åtgärder. Autonomin rör anomalidetektion och stöd till säkerhetscenter. Därtill har Dynatraces plattform en autonom lösning för exekveringsapplikationsskydd, vilket detekterar och blockerar angrepp som når till applikationen. Detta fungerar genom att skyddet antingen programmeras in i kodbasen eller läggs till som en agent vid exekvering av applikationen.

Privilegierad användarhantering (PAM) har autonomi i form av detektion och respons, genom en variant av användar- och enhetsbeteendeanalys för privilegierade användare.

NDR har i de flesta fall autonomi som rör intrångsdetektion och respons, medan detta alltid är sant för EDR. För EDR används även artificiell intelligens för att detektera angrepp som utnyttjar tidigare okända sårbarheter. Alla EDR-leverantörer i Gartners genomgång erbjuder även utökade funktioner med XDR och detta gäller även de flesta NDR-leverantörer. Med XDR tillkommer autonomi som rör användar- och enhetsbeteendeanalys samt stöd till säkerhetscenter.

Cyberfysiska systems skyddsplattformar (CPS PP) erbjuder framförallt autonomi i form av intrångsdetektion och respons. En del erbjuder även stöd för autonom hotjakt och stöd till säkerhetscenter.

Autonomi för moving target defence (AMTD) finns i en produkt. Gartner beskriver detta som teknik som utifrån detektion och med automatiserad respons gör oförutsägbara förändringar i ett it-system. Syftet är att förvirra och försvåra för angripare.

6 Diskussion

Detta kapitel diskuterar litteraturstudiens resultat, indelat per forskningsfråga. Dessutom beskrivs vilka metodmässiga begränsningar som finns med studien samt vilka framtida forskningsmöjligheter som finns.

6.1 Vilka tekniker används för att implementera aktiva skydd?

I forskningsartiklarna beskrivs framförallt tekniker som fokuserar på detektion. I undantagsfall förekommer även respons samt vilseledning i form av honungsfällor och autonom moving target defence. Detta är rimligt eftersom den största relevanta begränsningen hos människor förmodligen är att identifiera om något är fel, snarare än att agera vid fel. Det är en stor påfrestning för en människa att övervaka stora mängder nätverkstrafik och processer. Däremot är själva agerandet snarare ganska enkelt när rätt agerande avgjorts. Exempelvis är det enkelt för en typisk systemadministratör att stänga ner användarkonton, portar eller maskiner. Det finns dock inte alltid tillgång till en människa som kan agera och då behövs det även automatiska tekniker som agerar. Större ramverk för cyberförsvar och autonoma agenter som inkluderar flera tekniker är en nyare företeelse i forskningslitteraturen. I militär kontext förekommer det forskning och experiment med autonoma agenter för cyberförsvar.

Bland de kommersiella produkterna finns det många varianter av tekniker i de olika produkterna. Den autonomi som finns i produkterna är i form av intrångsdetektion, respons, avvikelsedetektion i form av användar- och enhetsbeteendeanalys samt diverse stöd till säkerhetscenter. Många av produkterna har funnits en längre tid och har med tiden anpassats till att ha utökad autonomi. Företagen beskriver dessa produkter som effektiva, men mognadsgraden av denna autonomi har inte utvärderats i denna rapport. Med tanke på populariserandet av AI de senaste åren är det också svårt att veta om företagen satt nya namn på gamla produkter av marknadsföringssyfte eller om de verkligen har en ny och effektiv autonom lösning.

6.2 Vilka för- och nackdelar har teknikerna?

Det varierar vilka hot som teknikerna i forskningsartiklarna ska skydda mot. Vanligast är tillgänglighetsangrepp (DoS) och skanning, men många andra hot förekommer också. Artiklarna saknar dock typiskt detaljer om hoten och även om de potentiella konsekvenserna. Det talar för att lösningarna är teknikfokuserade. Det är förväntat för forskningsartiklar, eftersom de normalt ska hållas generaliserbara snarare än fokusera på att bli konkreta. Detaljerna om hot är dock alltför knapphändiga för att kunna bedöma teknikernas säkerhetshöjande

effekter, med ett enstaka undantag. Vad gäller datasetens hot (t.ex. inspelade angrepp) är de ofta ganska naiva och enkla. Det är därmed en öppen fråga huruvida de autonoma teknikernas enkla träning skulle motstå mer avancerade hot i praktiken. Flera artiklar är någorlunda tydliga med att deras tekniker har svårt att upptäcka mer avancerade (eller annorlunda) hot. Exempelvis rör detta angrepp som liknar användarbeteende eller som utnyttjar okända sårbarheter.

I forskningsartiklarna utvärderar forskarna sina tekniker på ett övergripande sätt. Detta är naturligt eftersom forskning normalt hålls generell snarare än görs för enskilda system och miljöer. Men det gör det också svårt att bedöma i vilken utsträckning för- och nackdelar beskrivna i artiklarna skulle hålla i praktiken. Därtill är en risk att artiklarnas labbmiljöer i själva verket är speciella snarare än generella och att teknikerna anpassats för att fungera bra i just labbmiljöerna. För en AI-teknik kan detta att leda till överträning. Därtill är det svårt att säga något om teknikernas prestandakrav och resursförbrukning. Flera artiklar beskriver dock svagheter (eller styrkor) relaterat till resursförbrukning, med exempelvis stora mängder loggar och högdimensionella data.

Forskningsartiklarnas tekniker utvärderas ofta med mått som rör falska positiva, det vill säga i vilken utsträckning tekniken detekterar angrepp trots att angrepp inte sker. Falska positiva är förmodligen ett stort problem både för tekniker som föreslås i artiklar och för tekniker som används i kommersiella produkter. Varje förekomst av en falsk positiv kan ge resurskrävande utredningsarbete samt ge upphov till onödig respons som har negativa bieffekter. Åtgärder för att minska falska positiva leder tyvärr typiskt även till att det blir färre sanna positiva. Den lämpligaste balansen mellan låga falska positiva och höga sanna positiva beror på det specifika systemet. Kopplat till falska positiva tar artiklarna inte upp angripare som avsiktligt påverkar tekniken för att få det att agera med viss respons. I praktiken kan därför siffrorna för falska positiva bli ännu högre. Dessa avsiktliga angrepp mot autonoma tekniker finns bland annat beskrivna i forskning om fientlig maskininlärning (eng. adversarial machine learning). Mer om detta står i FOI-rapporterna av Axell m.fl. (2022) samt Kamrani m.fl. (2023).

Vad gäller produkternas för- och nackdelar presenteras produkterna typiskt som färdiga att installera och driftsätta, men utan att ange hur mycket möda och resurser det krävs att installera och driftsätta. Det är också i allmänhet svårt att jämföra olika kommersiella produkter, vilket även gäller för produkter som är inom samma produktkategori. Den låga jämförbarheten beror på att resultaten presentas på olika sätt och att det är svårt att få insyn i testprocesserna. För att få en klarare bild av produkternas förmågor krävs utförligare utvärdering. Exempelvis kan Mitres ATT&CK Evaluations (https://evals.mitre.org/) användas för att jämföra EDR-lösningar. Fördelen med en utvärdering av tredje part är att det sparar mycket tid. Samtidigt kanske resultaten från en sådan utvärdering inte är relevanta för den miljö eller hotbild som den egna organisationen ska använda systemet för. Det finns därmed värde i att bedriva egen utvärdering.

6.3 Vilka förutsättningar finns för att använda de identifierade teknikerna?

Bara enstaka forskningsartiklar har ett militärt fokus eller en motsvarande fältmiljö. För militära fältmiljöer är det framförallt två faktorer som är speciella: svårigheten i att få tag i bra data samt de potentiellt allvarliga konsekvenserna av felaktig respons. Dessa två faktorer diskuteras mer nedan, följt av en diskussion om integration och interoperabilitet.

Vad gäller svårigheten i att få tag i bra data finns det flera relevanta aspekter. En aspekt är konfidentialitet på så vis att det kan finnas begränsningar i vilka data som får användas för träning, test och vid skarp användning. En annan aspekt är att noderna i fältsystem kan röra på sig och det på ett oberäkneligt sätt samt utan anslutning till andra noder under längre eller återkommande perioder. Även när noderna har anslutning sker det typiskt på ett decentraliserat sätt nod till nod utan något centralt nav. Dessutom präglas vissa noder av låg bandbredd, vilket gör det svårare att exempelvis samla in data för att basera beslut på. Låg bandbredd kan även försvåra insamlandet av data för träning och testning. Därtill kan kalibrering behöva ske för vissa annars vanliga eller standardiserade kännetecken såsom tiden mellan paket. Om ett system avviker från något slags standardsystem kommer kännetecknen behöva uppdateras. Kanske kan systemkomponenter lånas ut till leverantörer av skyddstekniker som då kan träna och anpassa. Men detta beror på tilliten till leverantörerna.

Vad gäller de potentiellt allvarliga konsekvenserna av felaktig respons innebär det att det förmodligen finns lägre acceptans för autonomt agerande. I de forskningsartiklar som studerats i denna rapport saknas fokus på denna faktor. I annan forskning studeras dock detta, exempelvis utifrån möjligheten för människor att förstå de beslut som en AI tar. Ett sätt att möta detta är med förklarbar AI (eng. explainable AI, förkortat XAI) där människor kan få förklaringar i stil med att viss utdata mest baseras på en viss del av indata. Exempelvis kan beslutet att isolera en maskin förklaras med att det skett en ökning av utgående krypterad trafik från maskinen till en okänd IP-adress tillsammans med nekade åtkomster till känsliga filer. I praktiken kan dock sådana förklaringar bli alltför förenklade. AI-algoritmer som inte fungerar enligt enkla regelsystem kan inte helt förklaras med enkla regler. FOI-rapporten av Luotsinen m.fl. (2019) beskriver mer om möjliga sätt att förklara hur AI fungerar och specifikt då AI-algoritmernas beslut. Det finns också annan forskning om detta. I en forskningsartikel beskriver Hoffman m.fl. (2018) att människans förståelse av autonomi bland annat kan röra sig om att vilja veta vad systemet gjorde, varför det inte gjorde något annat, vad det hade gjort i en annan situation samt vad systemet kommer att göra härnäst. En brittisk studie om AI (Knack och Burke, 2024) intervjuade intressenter för att ta reda på acceptansen för olika grader av autonomi för olika typer av skydd. För detektion sågs den högsta graden av

autonomi som acceptabel. För isolering och vilseledning var graden något lägre och ännu lägre var den för avlägsnande. Detta stämmer också med vad Försvarsmaktens representanter uttryckte vid ett referensgruppsmöte i ett annat FMV-finansierat projekt. Vid mötet beskrevs att autonom detektion är betydligt mer önskvärt än autonomt agerande. I andra sammanhang talas det ibland inom olika militära domäner om olika former och grader av autonomi hos vapensystem. Sådana system kan exempelvis aktiveras av människor och sedan självständigt välja mål och agera mot målen, eller bara agera mot mål som valts av människor, eller istället agera självständigt men övervakas av människor. Liknande indelningar används även civilt och inom cyberdomänen. I verksamhetskritiska system finns rimligtvis ett större behov av att kunna förutsäga och kontrollera vad som sker i systemet i jämförelse med mindre kritiska system. Ju mer avancerade och komplexa skyddsteknikerna blir desto högre blir också kraven på förståelsen av hur skyddsteknikerna påverkar designen av systemet och vilka stödfunktioner som kan behövas för att hantera dem.

Arbetet med att införa och integrera teknikerna i ett specifikt system kan vara omfattande, både vid integration med andra skydd och vid integration med andra typer av systemkomponenter. Integrationsarbetet behöver förmodligen axlas av kunden snarare än leverantören av systemet. Forskningsartiklarna fokuserar inte på integrationssvårigheter och interoperabilitet. För kommersiella produkter beskriver de flesta av företagen att de egna produkterna är kompatibla med varandra och rekommenderar att en komplett produktsvit införskaffas. Från ett kompatibilitetsperspektiv är detta fördelaktigt, men det finns även nackdelar. Dels finns det risk för leverantörsinlåsning (eng. vendor lock-in), dels risk för att delsystemen i sviten inte är de individuellt bästa produkterna på marknaden. Förutom integration är det också svårt att bedöma svårigheten i att sköta drift av installerade produkter. De flesta tillverkare beskriver att installation av deras produkter är enkelt och problemfritt, men det presenteras inte fakta för att stärka dessa påståenden. Det är även svårt att få insikt i vilka krav på förutsättningar varje produkt har för att driftsätta tekniken.

Ytterligare en aspekt som kan påverka valet av skyddstekniker är godkännandeprocesser för system. I Försvarsmakten ska it-system generellt sett ackrediteras, vilket kan medföra att ett system förväntas ha en viss arkitektur och konfiguration efter att systemet driftsatts. Om ett system som ska ackrediteras inkluderar skyddstekniker som påverkar exempelvis dataflöden och nätverksarkitekturer behöver ackrediteringen ta höjd för sådana förändringar. Skyddstekniker som förändrar exempelvis nätverksarkitekturer och dataflöden kan också medföra att teknisk dokumentation inte överensstämmer med det faktiska systemet, vilken kan medföra merarbete vid vidareutveckling och incidenthantering.

6.4 Studiens begränsningar

I rapportens inledning beskrevs vissa avgränsningar som gjordes i studien. Framförallt avgränsades rapporten från skydd som är väldigt omogna, utgör motåtgärder i andras system eller inte är anpassade för militära fältmiljöer (och inte skulle kunna anpassas för sådana miljöer). Rapporten avgränsades också från etisk och juridisk analys av skyddens autonomi. Utöver dessa avgränsningar finns det andra begränsningar hos studien och dessa beskrivs nedan.

En begränsning ges av begreppsdjungeln. Redan begreppet *aktivt skydd* är svårdefinierat och det finns en lång rad andra överlappande och icke-stringent definierade begrepp i akademisk och kommersiell litteratur. Detta gjorde det svårt att hitta en bra söksträng som hittade allt av intresse. Därtill innebar begreppsförbistringen svårigheter att förstå litteraturen och ge en rättvis jämförelse mellan de olika teknikerna.

En annan begränsning ges av kvaliteten på litteraturen. Sökningarna efter forskningsartiklar gjordes i en databas (Scopus) som innehåller artiklar av både högre och lägre kvalitet. Rapportförfattarnas bedömning är att de artiklar som identifierats och inkluderats är av varierande kvalitet, vilket är enligt förväntan för en litteraturgenomgång. Exempelvis bedöms 11 av artiklarna vara knapphändiga med detaljer, 22 av acceptabel kvalitet och 7 av god kvalitet. För de allra flesta artiklar är det otydligt vilka forskningsfrågor eller hypoteser som skulle besvaras. Ett undantag är en artikel (ID 64) som utförde ett gediget experiment, med flera stora välkända dataset, samt presenterar både metod och data väl.

En tredje begränsning utgörs av studiens avvägning mellan djup och bredd. Som är typiskt för litteraturstudier är det inte möjligt att vara helt komplett. Studiens omfattning tillät inte ytterligare sökningar efter forskningsartiklar som refererar till de artiklar som valts ut. För de kommersiella lösningarna förlitar sig rapporten i huvudsak på Gartners identifierade produkter (vars oberoende gentemot tillverkarna är omtvistat). Informationen om produkterna är därmed inte baserad på oberoende vetenskapliga utvärderingar (såsom Mitres Evaluations). Å andra sidan är det troligt att förändringar sker snabbt på området varför snabba analyser kan vara att föredra framför mer djupgående sådana. Med anledning av de tidsmässiga begränsningarna gör rapporten inte en fullständig mappning mot ett lämpligt ramverk för att klassificera artiklarnas tekniker. Rapportförfattarna har därmed inte undersökt sådant som huruvida artiklarna tar med alla steg i en klassisk OODA-loop (observe, orient, decide, act), det vill säga: väljer loggkällor och hot, väger in systemkontexten, beslutar utifrån systemets syfte samt agerar med någon utvärderingsfunktion. Ett försök till klassificering mot ramverket Mitre D3fend gjordes dock för de kommersiella lösningarna.

6.5 Framtida forskningsmöjligheter

Det finns gott om möjligheter att utvidga den forskning som rapporterats här.

En möjlighet är att genomföra praktiska utvärderingar av olika kommersiella lösningar och av de tekniker som beskrivs inom forskningslitteraturen. I båda fallen kan mer djupgående förståelse för teknikerna uppnås, samtidigt som företags och forskares utfästelser kontrolleras. Det går också att fokusera på integrationstestning för att undersöka eventuella problem med installation och interoperabilitet. Det vore även intressant att genomföra en noggrannare analys av vilka tekniker som bäst passar i olika system och miljöer hos Försvarsmakten.

En annan möjlighet är att utforska arkitekturprinciper för att bygga it-system med aktivt skydd. Majoriteten av det som identifierats i studien handlar om enskilda tekniker, produkter eller i vissa fall produktsviter. Få källor belyser hur tekniker bör integreras i större system och vilka faktorer eller begränsningar som då bör beaktas. Det vore därför intressant att undersöka förmågan hos aktivt skydd i kompletta it-system.

En tredje möjlighet är att undersöka skillnader mellan de olika källorna. Det verkar finnas en diskrepans i skyddsteknikerna som förekommer i den akademiska litteraturstudien gentemot de som förekommer i kommersiella lösningar. De akademiska forskningsartiklar som studien identifierat handlar i stor utsträckning om intrångsdetektion eller intrångsprevention. De kommersiella lösningarna täcker också in aspekter såsom isolering, avlägsnande och återställande. Det vore intressant att utforska om denna diskrepans beror på begränsningar i den akademiska litteraturstudiens söktermer eller en faktisk brist på forskning. Kanske är skyddsteknikerna så väletablerade att det inte förekommer ny forskning på området. Å andra sidan ter detta sig osannolikt givet den snabba teknikutvecklingen i cyberdomänen.

7 Slutsatser

I forskningslitteraturen om aktiva cyberskydd beskrivs framförallt tekniker som fokuserar på detektion. Det förekommer också enstaka artiklar om respons samt vilseledning i form av honungsfällor och MTD. Det börjar också förekomma artiklar som beskriver större ramverk och autonoma agenter som kombinerar flera olika tekniker för detektion, beslut och respons. För de kommersiella produkterna finns autonomi i form av intrångsdetektion, respons, avvikelsedetektion och diverse stöd till SOC. Mognadsgraden av dessa delar av produkterna har dock inte bedömts i rapporten.

Teknikerna kan skydda mot många olika typer av hot och verkar fungera bra i olika miljöer enligt artiklarna, med exempelvis hög larmförmåga utan stora problem med falska positiva. Dessa utvärderingar är dock inte gjorda av oberoende parter. Därtill kan ingen teknik skydda mot alla hot och de flesta tekniker fokuserar på tillgänglighetsangrepp och skanning. Därtill beskriver litteraturen teknikerna på ett relativt övergripande sätt, vilket gör det svårt att säga huruvida de skulle fungera bra i praktiken. Riktiga hot, miljöer och situationer kan ställa krav som teknikerna inte kan möta.

Teknikerna är typiskt inriktade på miljöer som är ganska långt från den typiska militära fältmiljön, även om teknikerna kan anpassas för sådana miljöer. Därtill ställs krav på tillgång till data att träna och testa teknikerna med, vilket kan vara svårmött vid högre krav på konfidentialitet. En utmaning med teknikerna är – som med all autonomi – att den som förlitar sig på teknikerna inte alltid kan få kompletta förklaringar till rekommendationer och beteenden. Det finns också utmaningar med att få teknikerna att verka tillsammans med varandra och med andra komponenter.

Det finns gott om framtida forskningsmöjligheter på området. En möjlighet är att komplettera denna litteraturbaserade kunskapsöversikt med praktiska utvärderingar. En annan möjlighet vore att göra djupare analyser av skillnaderna mellan de kommersiella lösningarna och teknikerna från forskningslitteraturen.

8 Referenser

Nedan finns de allmänna referenser som använts samt de referenser som rör de forskningsartiklar som inkluderades i litteraturgenomgången.

8.1 Allmänna referenser

Axell E., Eliardsson P., Hägglund K., Brännström P., Svensson C. 2022. Adversarial Machine Learning in Wireless Communications. Basics and two examples. FOI-R--5427--SE. Totalförsvarets forskningsinstitut. https://foi.se/rapporter/rapportsammanfattning.html?reportNo=FOI-R--5427--SE

Bildsten C., Falkcrona J., Eidenskog D. 2021. Dynamiska honungsnätverk. Utmaningar och teknik för sömlös överflyttning av angrepp. FOI-R--5216--SE. Totalförsvarets forskningsinstitut. https://foi.se/rapporter/rapportsammanfattning.html?reportNo=FOI-R--5216--SE

Bildsten C., Karlzén H., Westerdahl L., Lundholm K., Eidenskog D., Karresand M. 2020. Förstudier inom informationssäkerhet. FOI-R--5068--SE. Totalförsvarets forskningsinstitut. https://foi.se/rapporter/rapportsammanfattning.html?reportNo=FOI-R--5068--SE

Center for Cyber & Homeland Security. 2016. Into the Gray Zone. The George Washington University.

Försvarsmakten. 2013. Försvarsmaktens redovisning av perspektivstudien 2013. Skrivelse. FM2013-276:1.

Försvarsmakten. 2022. Ett aktivt försvar – redo att agera. https://www.forsvarsmakten.se/sv/om-forsvarsmakten/darfor-finns-forsvarsmakten/forsvaret-av-sverige-i-dag-och-i-morgon/ett-aktivt-forsvar-redo-att-agera/ Hämtad 25 juni 2025.

Försvarsmakten. 2024a. Doktrinansats Cyberförsvar. 1.1.

Försvarsmakten. 2024b. DCO-konceptet. FM2024-9629:1

Försvarsmakten. u.å. Ordlista. Aktivt försvar. https://www.forsvarsmakten.se/sv/ordlista/#/word/aktivt-forsvar Hämtad 25 juni 2025.

Gartner. u.å. Magic Quadrant. https://www.gartner.com/en/research/magic-quadrant Hämtad 24 juni 2025.

Gartner. 2024a. Magic Quadrant for Endpoint Protection Platforms. Gartner.

Gartner. 2024b. Magic Quadrant Security Information and Event Management. Gartner.

Gartner. 2024c. Magic Quadrant for Observability Platforms. Gartner.

Gartner. 2024d. Magic Quadrant for Privileged Access Management. Gartner.

Gartner. 2025a. Magic Quadrant for CPS Protection Platforms. Gartner.

Gartner. 2025b. Magic Quadrant for Network Detection and Response. Gartner.

Hoffman R. R., Mueller S. T., Klein G., Litman J. 2018. Metrics for explainable AI: Challenges and prospects. arXiv preprint arXiv:1812.04608. https://arxiv.org/ftp/arxiv/papers/1812/1812.04608.pdf

Holm H., Bengtsson J., Löfvenberg J., Persson M., Sommestad T. 2014. Moving Target Defense. En kartläggning av forskningsbidrag. FOI-R--3942--SE. Totalförsvarets forskningsinstitut. https://foi.se/rapporter/rapportsammanfattning.html?reportNo=FOI-R--3942--SE

Kamrani F., Kanestad L., Luotsinen L., Pelzer B., Sabel J., Sandström V., Tegen A. 2023. Attacking and Deceiving Military AI Systems. FOI-R--5396--SE. Totalförsvarets forskningsinstitut. https://foi.se/rapporter/rapportsammanfattning.html?reportNo=FOI-R--5396--SE

Karlzén H. 2021. Honungsfällor. Att vilseleda och studera cyberangripare. FOI-R--5217--SE. Totalförsvarets forskningsinstitut. https://foi.se/rapporter/rapportsammanfattning.html?reportNo=FOI-R--5217--SE

Karlzén H., Gudmundson Hunstad A., Hyllienmark E., Rodhe I. 2020. Beteendebaserad biometrisk autentisering. FOI-D--0991--SE. Totalförsvarets forskningsinstitut.

Karlzén H., Sommestad T. 2023. Automatic incident response solutions: a review of proposed solutions' input and output. ARES 2023. https://dl.acm.org/doi/pdf/10.1145/3600160.3605066

Karlzén H., Valassi C. 2022. Detektion av illasinnad exfiltrering av data via nätverk. En systematisk litteraturgenomgång. FOI-R--5376--SE. Totalförsvarets forskningsinstitut. https://foi.se/rapporter/rapportsammanfattning.html?reportNo=FOI-R--5376--SE

Knack A., Burke A. 2024. Autonomous Cyber Defence - Authorised bounds for autonomous agents. CETaS Briefing paper.

Luotsinen L.J., Oskarsson D., Svenmarck P., Wickenberg Bolin U. 2019. Explainable Artificial Intelligence: Exploring XAI Techniques in Military Deep Learning Applications. FOI-R--4849--SE. https://foi.se/rapporter/rapportsammanfattning.html?reportNo=FOI-R--4849--SE

Mitre. 2025. D3FEND. A knowledge graph of cybersecurity countermeasures. 1.1.0. https://d3fend.mitre.org Hämtad 15 juli 2025.

Pollock N., Williams R. 2009. The sociology of a market analysis tool: How industry analysts sort vendors and organize markets. Information and Organization, 19(2), 129-151.

Rodhe I., Westring E., Karlzén H. 2014. Self-healing and self-protecting software. Scanning the research frontier. FOI-R--3836--SE. https://foi.se/rapporter/rapportsammanfattning.html?reportNo=FOI-R--3836--SE

Svenska Akademien. 2021a. SO. skydd. https://svenska.se/tre/?sok=skydd&pz=1

Svenska Akademien. 2021b. SO. försvar. https://svenska.se/tre/?sok=f%C3%B6rsvar&pz=1

Zouave E. 2020. De svenska definitionerna inom aktivt cyberförsvar. FOI-D--0981--SE. Totalförsvarets forskningsinstitut.

8.2 Litteraturgenomgångens forskningsartiklar

ID	Referens
4	Lee D., Kim D., Ahn M.K., Lee S. 2024. Bayesian Stackelberg game approach for cyber mission impact assessment. ICT Express, 10(2). DOI: 10.1016/j.icte.2023.11.003
7	Couillard M., Hale B. 2024. DRACO: Production Network Deployment and Evaluation of Deceptive Defense As-a-Service. Proceedings - 2024 IEEE International Conference on Big Data, BigData 2024. DOI: 10.1109/BigData62323.2024.10825309
8	Demertzis K., Tziritas N., Kikiras P., Sanchez S.L., Iliadis L. 2019. The next generation cognitive security operations center: Adaptive analytic lambda architecture for efficient defense against adversarial attacks. Big Data and Cognitive Computing, 3(1). DOI: 10.3390/bdcc3010006
10	Shen S., Cai C., Shen Y., Wu X., Ke W., Yu S. 2024. MFGD3QN: Enhancing Edge Intelligence Defense Against DDoS With Mean-Field Games and Dueling Double Deep Q-Network. IEEE Internet of Things Journal, 11(13). DOI: 10.1109/JIOT.2024.3387090
11	Feng C., Von Der Assen J., Huertas Celdrán A., Näf S., Bovet G., Stiller B. 2023. FeDef: A Federated Defense Framework Using Cooperative Moving Target Defense. 2023 8th International Conference on Smart and Sustainable Technologies, SpliTech 2023. DOI: 10.23919/SpliTech58164.2023.10193681
25	Feng W., Vyas S., Li T. 2025. Autonomous Cyber Defence by Quantum-Inspired Deep Reinforcement Learning. International

	Conference on Information Systems Security and Privacy, 2. DOI: 10.5220/0013151800003899
29	Liu Y., Zhao J., Zhang G., Xing C. 2021. NetObfu: A lightweight and efficient network topology obfuscation defense scheme. Computers and Security, 110. DOI: 10.1016/j.cose.2021.102447
42	Yurekten O., Demirci M. 2021. Citadel: Cyber threat intelligence assisted defense system for software-defined networks. Computer Networks, 191. DOI: 10.1016/j.comnet.2021.108013
53	Seo S., Moon H., Lee S., Kim D., Lee J., Kim B., Lee W., Kim D. 2023. D3GF: A Study on Optimal Defense Performance Evaluation of Drone-Type Moving Target Defense Through Game Theory. IEEE Access, 11. DOI: 10.1109/ACCESS.2023.3278744
54	Tang J., Chen M., Chen H., Zhao S., Huang Y. 2023. A new dynamic security defense system based on TCP_REPAIR and deep learning. Journal of Cloud Computing, 12(1). DOI: 10.1186/s13677-022-00379-2
56	Raio S., Corder K., Parker T.W., Shearer G.G., Edwards J.S., Thogaripally M.R., Park S.J., Nelson F.F. 2023. Reinforcement Learning as a Path to Autonomous Intelligent Cyber-Defense Agents in Vehicle Platforms. Applied Sciences (Switzerland), 13(21). DOI: 10.3390/app132111621
64	Yi H., Zhang S., An D., Liu Z. 2024. PatchesNet: PatchTST-based multi-scale network security situation prediction. Knowledge-Based Systems, 299. DOI: 10.1016/j.knosys.2024.112037
88	Loevenich J., Adler E., Hürten T., Lopes R.R.F. 2025. Design and evaluation of an Autonomous Cyber Defence agent using DRL and an augmented LLM. Computer Networks, 262. DOI: 10.1016/j.comnet.2025.111162
113	Jaber A. 2024. Transforming Cybersecurity Dynamics: Enhanced Self-Play Reinforcement Learning in Intrusion Detection and Prevention System. SysCon 2024 - 18th Annual IEEE International Systems Conference, Proceedings. DOI: 10.1109/SysCon61195.2024.10553626
124	Bradley T., Watkins L., Alhajjar E. 2023. Autonomic Cyber Security Enhanced with Survival Analysis (ACSeSA). Proceedings - Conference on Local Computer Networks, LCN. DOI: 10.1109/LCN58197.2023.10223332
140	Islam M.M., Duan Q., Al-Shaer E. 2019. Specification-driven moving target defense synthesis. Proceedings of the ACM Conference on Computer and Communications Security. DOI: 10.1145/3338468.3356830
153	Mi Y., Mohaisen D., Wang A. 2022. AutoDefense: Reinforcement Learning Based Autoreactive Defense Against Network Attacks. 2022 IEEE Conference on Communications and Network Security, CNS 2022. DOI: 10.1109/CNS56114.2022.9947232

163	Islam M.M., Al-Shaer E. 2020. Active deception framework: An extensible development environment for adaptive cyber deception. Proceedings - 2020 IEEE Secure Development, SecDev 2020. DOI: 10.1109/SecDev45635.2020.00023
173	Muhati E., Rawat D.B. 2021. Asynchronous Advantage Actor-Critic (A3C) Learning for Cognitive Network Security. Proceedings - 2021 3rd IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications, TPS-ISA 2021. DOI: 10.1109/TPSISA52974.2021.00012
195	Wang X., Shi L., Cao C., Wu W., Zhao Z., Wang Y., Wang K. 2024. Game analysis and decision making optimization of evolutionary dynamic honeypot. Computers and Electrical Engineering, 119. DOI: 10.1016/j.compeleceng.2024.109534
218	Chatterjee S., Shaw V., Das R. 2024. Multi-stage intrusion detection system aided by grey wolf optimization algorithm. Cluster Computing, 27(3). DOI: 10.1007/s10586-023-04179-4
224	Odeh A., Abu Taleb A. 2024. Robust Network Security: A Deep Learning Approach to Intrusion Detection in IoT. Computers, Materials and Continua, 81(3). DOI: 10.32604/cmc.2024.058052
229	Al-Ofeishat H.A., Alkasassbeh J.S., Awajan A., Alazab M. 2025. Analysis and Comparison of Raw Network Packet Datasets Using Machine Learning Classification and Grey Wolf Optimization. International Journal of Advances in Soft Computing and its Applications, 17(1). DOI: 10.15849/IJASCA.250330.13
232	Nguyen V.Q., Ngo L.T., Nguyen V.H., Nguyen L.M., Le-Khac N.-A. 2024. A Deep Metric Learning Approach for Cyber Reconnaissance Detection. 1st International Conference on Cryptography and Information Security, VCRIS 2024 - Proceedings. DOI: 10.1109/VCRIS63677.2024.10813453
233	Park D., Min B., Lim S., Kim B. 2025. ATIRS: Towards Adaptive Threat Analysis with Intelligent Log Summarization and Response Recommendation. Electronics (Switzerland), 14(7). DOI: 10.3390/electronics14071289
278	Tariq U., Ullah I., Yousuf Uddin M., Kwon S.J. 2022. An Effective Self-Configurable Ransomware Prevention Technique for IoMT. Sensors, 22(21). DOI: 10.3390/s22218516
281	Myers J., Babun L., Yao E., Helble S., Allen P. 2019. MAD-IoT: Memory anomaly detection for the internet of things. 2019 IEEE Globecom Workshops, GC Wkshps 2019 - Proceedings. DOI: 10.1109/GCWkshps45667.2019.9024539
349	Abdalgawad N., Sajun A., Kaddoura Y., Zualkernan I.A., Aloul F. 2022. Generative Deep Learning to Detect Cyberattacks for the IoT-23 Dataset. IEEE Access, 10. DOI: 10.1109/ACCESS.2021.3140015

352	Hatzivasilis G., Soultatos O., Chatziadam P., Fysarakis K., Askoxylakis I., Ioannidis S., Alexandris G., Katos V., Spanoudakis G. 2021. WARDOG: Awareness Detection Watchdog for Botnet Infection on the Host Device. IEEE Transactions on Sustainable Computing, 6(1). DOI: 10.1109/TSUSC.2019.2914917
355	Nespoli P., Marmol F.G., Vidal J.M. 2021. A Bio-Inspired Reaction against Cyberattacks: AIS-Powered Optimal Countermeasures Selection. IEEE Access, 9. DOI: 10.1109/ACCESS.2021.3074021
367	Skopik F., Landauer M., Wurzenberger M., Vormayr G., Milosevic J., Fabini J., Prüggler W., Kruschitz O., Widmann B., Truckenthanner K., Rass S., Simmer M., Zauner C. 2020. synERGY: Cross-correlation of operational and contextual data to timely detect and mitigate attacks to cyber-physical systems. Journal of Information Security and Applications, 54. DOI: 10.1016/j.jisa.2020.102544
377	Shaaban G., Fourati H., Kibangou A., Prieur C. 2025. Active Defense Strategy in Cyber-Physical Systems: Misleading Unauthorized Observers. IEEE Transactions on Control of Network Systems. DOI: 10.1109/TCNS.2025.3570931
380	Enoch S.Y., Moon C.Y., Lee D., Ahn M.K., Kim D.S. 2022. A practical framework for cyber defense generation, enforcement and evaluation. Computer Networks, 208. DOI: 10.1016/j.comnet.2022.108878
386	Chai X., Wang Y., Yan C., Zhao Y., Chen W., Wang X. 2020. DQ-MOTAG: Deep reinforcement learning-based moving target defense against DDoS attacks. Proceedings - 2020 IEEE 5th International Conference on Data Science in Cyberspace, DSC 2020. DOI: 10.1109/DSC50466.2020.00065
387	Singh A.V., Rathbun E., Graham E., Oakley L., Boboila S., Chin P., Oprea A. 2025. Hierarchical Multi-agent Reinforcement Learning for Cyber Network Defense: Extended Abstract. Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS. DOI: 10.48550/arXiv.2410.17351
389	Husák M., Laštovička M., Tovarnak D. 2021. System for Continuous Collection of Contextual Information for Network Security Management and Incident Handling. ACM International Conference Proceeding Series. DOI: 10.1145/3465481.3470037
392	Xu S., Shi Y., Shi L., Zhang H. 2025. Efficient network defense policies via GNN-enhanced reinforcement learning. Journal of Supercomputing, 81(8). DOI: 10.1007/s11227-025-07431-3
394	Loevenich J.F., Adler E., Mercier R., Velazquez A., Lopes R.R.F. 2024. Design of an Autonomous Cyber Defence Agent using Hybrid AI models. 2024 International Conference on Military Communication and Information Systems, ICMCIS 2024. DOI: 10.1109/ICMCIS61231.2024.10540988

427	Saeed M.M. 2025. An AI-Driven Cybersecurity Framework for IoT: Integrating LSTM-Based Anomaly Detection, Reinforcement Learning, and Post-Quantum Encryption. IEEE Access, 13. DOI: 10.1109/ACCESS.2025.3576506
433	Yaman F., Eskridge T., Adler A., Atighetchi M., Simidchieva B.I., Jeter S., Cassetti J., DeMatteis J. 2020. An autonomous resiliency toolkit for cyber defense platforms. ICAART 2020 - Proceedings of the 12th International Conference on Agents and Artificial Intelligence, 2. DOI: 10.5220/0009142702400248
453	Saurabh K., Singh S., Vyas R., Vyas O.P., Khondoker R. 2022. MLAPS: A Machine Learning based Second Line of Defense for Attack Prevention in IoT Network. INDICON 2022 - 2022 IEEE 19th India Council International Conference. DOI: 10.1109/INDICON56171.2022.10039777
461	Oroian D., Bolboaca R., Roman A.-S., Dobrota V. 2024. Network Intrusion Detection System Using Anomaly Detection Techniques. Proceedings - 2024 IEEE 20th International Conference on Intelligent Computer Communication and Processing Conference, ICCP 2024. DOI: 10.1109/ICCP63557.2024.10793023
466	Liu Z., Su N., Qin Y., Lu J., Li X. 2020. A Deep Random Forest Model on Spark for Network Intrusion Detection. Mobile Information Systems, 2020. DOI: 10.1155/2020/6633252
472	Anley M.B., Genovese A., Agostinello D., Piuri V. 2024. Robust DDoS attack detection with adaptive transfer learning. Computers and Security, 144. DOI: 10.1016/j.cose.2024.103962
474	Samaddar A., Potteiger N., Koutsoukos X. 2025. Out-of-Distribution Detection for Neurosymbolic Autonomous Cyber Agents. 2025 IEEE 4th International Conference on AI in Cybersecurity, ICAIC 2025. DOI: 10.1109/ICAIC63015.2025.10849024
480	Das B.C., Sartaz M.S., Reza S.A., Hossain A., Nasiruddin M.D., Bishnu K.K., Sultana K.S., Shatyi S.S., Khan M.D.A., Abed J. 2025. AI-Driven Cybersecurity Threat Detection: Building Resilient Defense Systems Using Predictive Analytics. International Journal of Basic and Applied Sciences, 14(4). DOI: 10.14419/hysdg957
495	Alavizadeh H., Alavizadeh H., Jang-Jaccard J. 2022. Deep Q-Learning Based Reinforcement Learning Approach for Network Intrusion Detection. Computers, 11(3). DOI: 10.3390/computers11030041
496	Ansah P., Tetarave S.K., Kalaimannan E., Dash B.B., John C. 2023. Enhancing Network Security Through Proactive Anomaly Detection: A Comparative Study of Auto-Encoder Models and K-Nearest Neighbours Algorithm. 2023 3rd Intelligent Cybersecurity Conference, ICSC 2023. DOI: 10.1109/ICSC60084.2023.10349990

540	Pu C., Lim S., Chae J., Jung B. 2019. Active detection in mitigating routing misbehavior for MANETs. Wireless Networks, 25(4). DOI: 10.1007/s11276-017-1621-z
548	Aminanto M.E., Zhu L., Ban T., Isawa R., Takahashi T., Inoue D. 2019. Automated Threat-Alert Screening for Battling Alert Fatigue with Temporal Isolation Forest. 2019 17th International Conference on Privacy, Security and Trust, PST 2019 - Proceedings. DOI: 10.1109/PST47121.2019.8949029
555	Blakely B., Billings H., Evans N., Landry A., Domingo A. 2023. Evaluation of an Autonomous Intelligent Cyberdefense Agent at NATO Cyber Coalition Exercise 2022. Proceedings of SPIE - The International Society for Optical Engineering, 12542. DOI: 10.1117/12.2662959
573	Thanigaivel G., Yeswanth A. 2025. CyberSentinel: Machine Learning-Driven Attack Severity Prediction & Threat Intelligence Using UNSW_NB15 Dataset. 2nd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering, RMKMATE 2025. DOI: 10.1109/RMKMATE64874.2025.11042779
599	Velazquez A., Lopes R.R.F., Bécue A., Loevenich J.F., Rettore P.H.L., Wrona K. 2023. Autonomous Cyber Defense Agents for NATO: Threat Analysis, Design, and Experimentation. MILCOM 2023 - 2023 IEEE Military Communications Conference: Communications Supporting Military Operations in a Contested Environment. DOI: 10.1109/MILCOM58377.2023.10356321
626	Schabinger R.M., Carlin C., Mullin J., Bierbrauer D.A., Nack E.A., Pavlik J.A., Wei A.V., Bastian N.D., Ahiskali M.B. 2024. Dynamic Reinforcement Learning for Network Defense: Botnet Detection and Eradication. Proceedings of SPIE - The International Society for Optical Engineering, 13051. DOI: 10.1117/12.3012783
684	Steverson K., Carlin C., Mullin J., Ahiskali M. 2021. Cyber Intrusion Detection using Natural Language Processing on Windows Event Logs. 2021 International Conference on Military Communication and Information Systems, ICMCIS 2021. DOI: 10.1109/ICMCIS52405.2021.9486307

ISSN 1650-1942