Myndigheten för samhällsskydd och beredskap

FORSKNING/STUDIE

CERCES2 – Center för resilienta kritiska infrastrukturer

Populärvetenskaplig slutrapport

CERCES2 – Center för resilienta kritiska infrastrukturer

Tidsperiod: september 2020 – oktober 2024
Utförare: KTH
Ansvarig: Henrik Sandberg, Ragnar Thobaben, György Dán, Mads Dam

Kort sammanfattning: Forskare vid KTH har under fyra års tid arbetat med att utveckla nya tekniklösningar för att stödja säkerhetsarbetet i kritisk infrastruktur. Projektet fullföljer det arbete som inleddes under CERCES-projektet.

MSB:s Kontaktpersoner: Joachim Elevant, Erik Sundström, 010-240 53 71

Foto omslag: CERCES2 logo (Design: Brinton Seashore-Ludlow)
Text: Henrik Sandberg, Ragnar Thobaben, György Dán, Mads Dam

Publikationsnummer: MSB2505 – december 2024

MSB har beställt och finansierat genomförandet av denna forskningsrapport (alt. studierapport). Författarna är ensamma ansvariga för rapportens innehåll.

Förord

Under de senaste åren har det genomförts cyberattacker mot industriell infrastruktur med stora skador och kostnader som följd. Detta har kunnat ske, eftersom mycket av den teknik som används i samhällsviktiga verksamheter har digitaliserats. Digitalisering, och ökad användning av it, ger många fördelar men leder också till att nya risker uppkommer. En sådan ny risk är att komplexiteten i form av beroenden ökar. Även exponeringen av system ökar när information och applikationer distribueras till fler enheter, externa organisationer och kan bli åtkomliga och därmed sårbara.

Elförsörjning, vattenförsörjning och transporter är exempel på samhällsviktiga verksamheter som idag beroende av it för att bedriva sin verksamhet. För att styra och övervaka den infrastruktur och de processer som behövs för att dessa verksamheter ska fungera används industriella informations- och styrsystem, även kallade cyberfysiska system. Avbrott i dessa system kan leda till störningar i samhällsviktiga funktioner, vilket kan leda till allvarliga konsekvenser för samhället.

Kriget i Ukraina har visat att civila samhällsviktiga verksamheter är legitima mål. Angrepp mot cyberfysiska system riskerar under höjd beredskap eller krig att minska det civila försvarets förmåga att stödja det militära försvaret.

Mot bakgrund av de cyberfysiska systemens betydelse för samhället har MSB satsat på utveckling av kunskap om systemens sårbarheter och deras lösningar. Forskningsprojektet CERCES-2, en fortsättning på projektet CERCES, har arbetat inom fyra områden och har gjort stora framsteg i vilka säkerhetsåtgärder som måste stärkas för att skydda våra viktiga industriella informations- och styrsystem.

Lars-Göran Emanuelson

Enhetschef, Avdelningen för cybersäkerhet och säkra kommunikationer

Innehåll

INLEDNING **5**

EN FÖRBÄTTRAD FÖRSTÅELSE FÖR ATTACKSTRATEGIER MOT TRÅDLÖSA AUTENTISERINGSMETODER **6**

KOMMUNIKATIONS- OCH BERÄKNINGSINFRASTRUKTURER **8**

RÖRLIGA FÖRSVAR OCH SPELTEORI **9**

SÄKERHET I INBYGGDA SYSTEM **11**

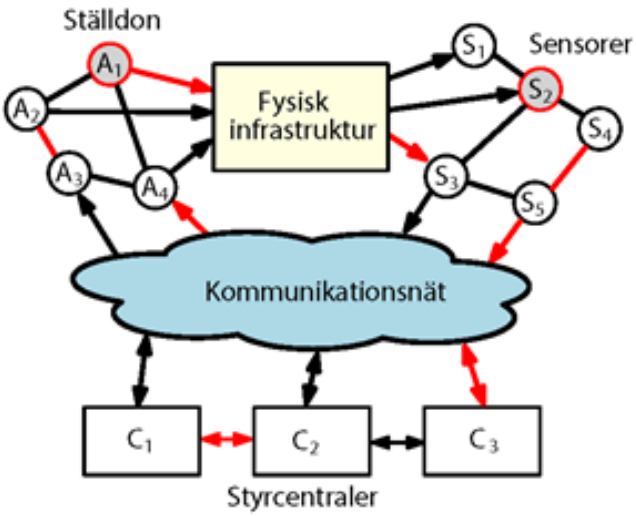
FÖRDJUPAD LÄSNING **13**

Inledning

Samhällets kritiska infrastrukturer, som elnät och andra distributionsnät, övervakas och regleras med hjälp av industriella informations- och styrsystem, numera även kallade OT-system (Operational Technology System) i kontrast till IT-system. Om dess komponenter utsätts för cyberattacker riskerar det att leda till stora samhällsstörningar och fysisk skada. Forskare inom CERCES2-projektet har under fyra års tid utvecklat nya säkerhetslösningar för denna speciella miljö, som en uppföljning till det tidigare CERCES-projektet.

Ett oroande hot är att storskaliga industriella informations- och styrsystem, som används för övervakning och reglering, utsätts för cyberattacker (se figur 1 och även [1]). I värsta fall kan sådana attacker leda till att kritisk infrastruktur såsom elnät och trafiksystem fallerar med stora samhällsskador som följd. Ett exempel på detta kommer från Ukraina där 30 ställverk i elnätet stängdes ner p.g.a. en cyberattack den 23 december 2015. Detta fick till följd att över 200 000 kunder förlorade tillgång till elkraft [2]. Under 2016 och 2022 genomfördes än mer avancerade attacker mot närliggande elnät. En särskild mjukvara kallad Industroyer hade utvecklats för att skapa elavbrott [3]. Ett flertal liknande attackverktyg, med namn som Stuxnet, Duqu, BlackEnergy och Triton, särskilt riktade mot industriella informations- och styrsystem, har rapporterats om i media.

Figur 1. Ett industriellt informations- och styrsystem, ett s.k. OT-system, för kritisk fysisk infrastruktur kan ha många komponenter och kommunikationskanaler som är känsliga för cyberangrepp (indikerade i rött). I CERCES2 utvecklades nya metoder för att skydda sådana känsliga systemelement



I denna rapport beskrivs några av de resultat som togs fram i CERCES2-projektet som under åren 2020–2024 studerad och utvecklade nya säkerhetslösningar för dessa så kallade cyberfysiska hot. Projektet byggde vidare på resultat från det tidigare CERCES-projektet och har varit aktivt i fyra olika teknikområden, vilket avspeglar att säkerhetshoten måste förstås och bemötas på flera nivåer parallellt. Nedan följer en kort sammanställning av fyra olika arbeten i dessa teknikområden.

En förbättrad förståelse för attackstrategier mot trådlösa autentiseringsmetoder

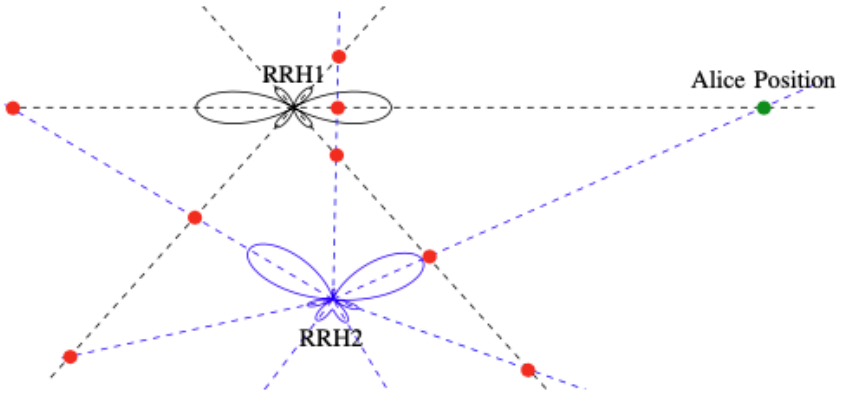
Trådlös kommunikation blir en allt viktigare teknik inom industriella informations- och styrsystem. Stora delar av kommunikationsnätet i figur 1 kan vara trådlöst vilket förstås ger upphov till nya sårbarhet. Till exempel kan sändare skicka falska meddelanden och störa annan trafik, vilket vi i CERCES2 studerat genom analys av olika möjliga metoder för att identifiera sändaren.

Metoder som utnyttjar den trådlösa radiokanalen för autentisering (t.ex. [4–7]) använder egenskaper hos kanalen som antingen kan knytas till sändarens position eller sändarens hårdvara. Om man använder ett flerantennsystem så blir överföringen riktad; det vill säga likt en ficklampa kan man skicka elektromagnetiska vågor åt ett bestämt håll, och likt en tratt kan en mottagare ”lyssna” åt ett bestämt håll. Denna teknik kallas även för ”beamforming” och kopplar radiokanalen tydligt till sändarens position.

Ett flerantennsystem som är kopplat till endast en mottagare kan bara avgöra från vilket håll en mottagen överföring har skickats. Det gör systemet sårbar mot attacker där en attackerare försöker att utge sig som en legitim sändare i systemet; en attackerare som står i linje med en legitim sändare eller på en speglad position har en stor chans att lyckas med sin attack.

I vår forskning [4] har vi visat att man kan signifikant minska antalet gynnsamma attackpositioner igenom att kombinera fler flerantennsystem som är placerade i olika positioner. Denna typ av distribuerade flerantennsystem kan exakt lokalisera sändarens position och analytisk bestämma vilka attackpositioner överhuvudtaget är relevanta. En illustration visas i figur 2 där två flerantennmottagare (RRH1 and RRH2) kombineras för att lokalisera den legitima sändaren Alice. Om systemet endast använder RRH1 för att lokalisera Alice så skulle varje punkt längs de svarta randiga linjerna utgöra en gynnsam attackposition. Genom att kombinera båda mottagare reduceras mängden av attackpositionerna till de röda punkterna där linjerna möts.

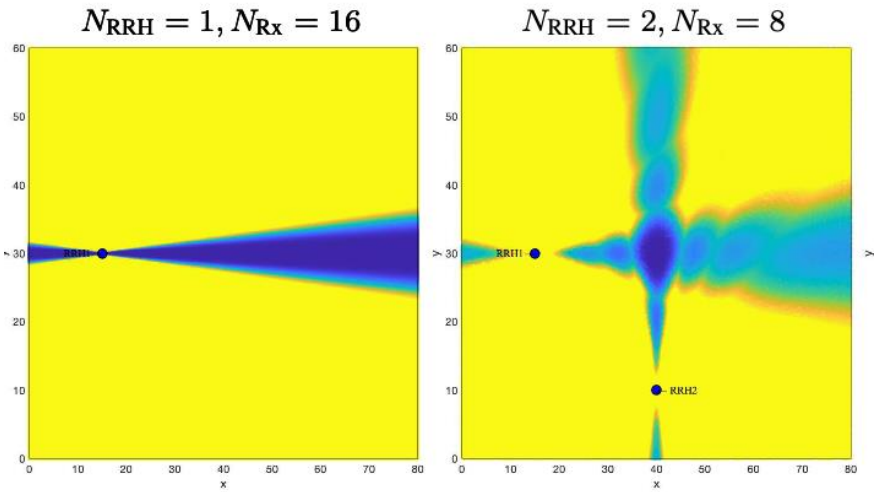
Figur 2. Distribuerad flerantennsystem med två mottagare (RRH1 och RRH2) som är riktade mot den legitima användaren Alice. De randiga linjerna indikerar riktningar därifrån signaler kan tas emot. Positioner där linjerna möts (markerade med röda punkter) är de enda möjliga attackpositioner där en attackerare har en ökad chans att vara framgångsrik



Figur 3 illustrerar hur sannolikheten för en framgångsrik attack påverkas. Regioner med hög sannolikhet för framgång (det vill säga nära 1) är markerade med mörkblå färg. Ljusblå markerar områden där sannolikheten för framgång ligger runt 10^{-4}, och sannolikheten för framgång i de gula områden är 10^{-16} eller lägre. Resultatet visar alltså att två mottagare gör det möjligt att effektivt minskar sannolikheten för en framgångsrik attack så länge attackeraren inte står nära den legitima sändaren.

I vår analys [4] har vi använt denna och liknande tekniker och resultat för att bestämma attackpositionerna och -parametrarna som leder till den största sannolikheten för en framgångsrik attack. Dessa parametrar har vi sedan använt i vår analys av systemets prestanda [5, 6] vilket lett till en kvantifiering av prestandanivån som alltid kan garanteras, även om systemet är under attack. Denna typ av resultat kan alltså fungera som en viktig säkerhetsgaranti för trådlösa autentiseringsmetoder.

Figur 3. Illustration av sannolikheten för en missad detektion av attackeraren för distribuerade flerantennsystem med totalt 16 antenner. Till vänster: en mottagare med 16 antenner; till höger: två mottagare med 8 antenner. Mörkblå indikerar regioner med hög sannolikhet för en missad attackdetektion



Kommunikations- och beräkningsinfrastrukturer

Attacker kan även genomföras i trådbundna kommunikationsnät och i CERCES2 har vi särskilt studerat sådana nät för reglering av elnät. Tidsynkronisering över nätet är en alltmer använd teknik för att mäta tillståndet i elnätet i realtid och givetvis öppnar det upp för nya typer av angrepp som vi analyserat.

Vi har fokuserat på att analysera sårbarheter mot attacker i kommunikationsinfrastrukturen i smarta elnät. Å ena sidan har vi utvecklat maskininlärningsalgoritmer för att förstå sårbarheter i distribuerade algoritmer som används för att optimera elnätets användande [8]. Vi använde automatisk generationsreglering som ett exempel, och har tränat en maskinlärningsalgoritm för att vilseleda algoritmen som utför regleringen för att orsaka instabilitet i elnätet. Våra resultat visar att en angripare som har tillgång till tillräcklig information om ett elnät kan lära sig icke detekterbara attacker mot elnätet utan att ha en särskilt djup förståelse av fysikens lagar. Å andra sidan har vi utvecklat nya algoritmer för att detektera och identifiera attacker mot tidssynkronisering i smarta elnät baserat på en tolkning av mätdata som en så kallad grafsignal [9]. Algoritmen som har utvecklats kan köras i realtid och har hög precision både för attackdetektering och för lokalisering.

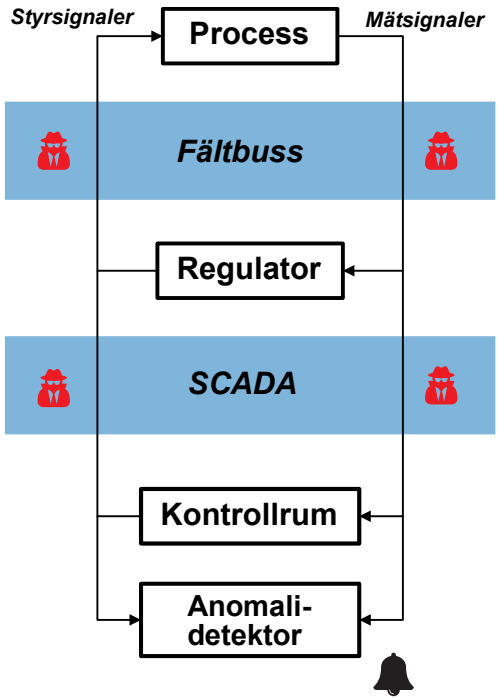
Vidare har nya typer av angrepp mot tidssynkroniseringsprotokoll identifierats, till exempel genom tillgång till s.k. switchar i kommunikationsnätet. Metoder har utvecklats för detektion av och försvar mot sådana angrepp, genom statistisk analys av mätdata och tidssynkroniseringssignaler, med tillämpning inom synkronmätningar i elnät. Vidare har resilient allokering av så kallade virtuella regulatorer för processtyrning studerats. Artiklarna [8] och [9] beskriver arbetet i detalj och inkluderar referenser till ytterligare material.

Rörliga försvar och spelteori

Under CERCES utvecklades flera metoder för att dimensionera industriella informations- och styrsystem så att möjliga skador i olika attacksituationer begränsas. Detta var ett preventivt försvar. Under CERCES2 fokuserade arbetet i stället på reaktivt försvar, där vi minimerar skador under pågående anfall. Särskilt tittade vi på så kallade rörliga försvar (moving-target defense).

En särskild typ av farlig och avancerad attack i cyberfysisk miljö är så kallade smygattacker (stealth attacks). I figur 4 visas schematiskt ett industriellt styrsystem med möjliga attackpunkter. Dessa attacker manipulerar data i systemet så att anomalidetektorn inte larmar operatören, eller fördröjer larmet så att stora skador kan uppkomma. En smygattack kan bara genomföras av en angripare med stor kännedom om systemets alla komponenter och delsystem. Till exempel kan angriparen passivt samla in data från SCADA-nätverket under en längre tid för att bygga upp modeller av systemet. Modellerna används sedan för att generera anfall som anomalidetektorn har svårt att upptäcka. Vi illustrerade detta experimentellt i [10].

Figur 4. Blockschemat illustrerar hur olika komponenter i ett förenklat industriellt reglersystem kommunicerar och hur potentiella cyberattacker (i rött) kan störa signalvägar och enheter på olika nivåer. Mätsensorn i en process kan till exempel manipuleras av en angripare så att regulatorn eller kontrollrummet applicerar skadliga styrsignaler. En anomalidetektor ska varna för sådana manipulationer men kan luras av så kallade smygattacker



En lockande försvarsstrategi mot smygattacker är att då och då ändra system-konfigurationerna så att anfallet avslöjas eller fördröjs. Samtidigt kan det vara komplicerat och riskfyllt att ändra i ett tidskritiskt cyberfysiskt system som fungerar väl, särskilt som man ofta inte förväntar sig att en smygangripare befinner sig i systemet. Man hamnar således i en situation där flera risker måste vägas mot varandra. Vi har en målkonflikt och är intresserade av att undersöka bästa möjliga försvarsstrategier.

För att studera detta problem utvecklades under CERCES2 verktyg som med spelteori karakteriserar optimala, slumpvisa försvarsstrategier för att ändra inställningar i delsystem. Särskilt identifierade vi anomalidetektorns tröskelvärde som en lämplig parameter att ändra på. Tröskelvärdet avgör vilka signalnivåer som ska bedömas som normala och vilka som ska resultera i alarm. Eftersom det alltid förekommer slumpvisa störningar kommer ett lågt tröskelvärde leda till många falsklarm och att operatörers förtroende för detektorn minskar över tid. Ett stort tröskelvärde kommer i stället leda till ett stort utrymme för en smygattack.

I artikeln [11] modellerade vi det cyberfysiska systemet som ett spel mellan en smygangripare och en operatör, med motstridiga krav på systemet. Den optimala försvarsstrategin kan beräknas som en s.k. Bayesiansk Nashjämvikt. Denna visar att *operatören oftast ska använda ett högt tröskelvärde,* med få falsklarm som följd, men att operatören då och då *slumpvis kraftigt ska sänka tröskelvärdet*. Detta hot om att tröskelvärdet kan sänkas kraftigt begränsar den skada smygangriparen kan uppnå. Försvarsstrategin ger operatören en möjlighet att, i någon mening, kringgå målkonflikten och få både ett lågt antal falsklarm i genomsnitt och en liten skada från smyganfall.

Säkerhet i inbyggda system

Inbyggda processorer hittar man i flera olika nivåer i industriella informations- och styrsystem. Speciellt finner man dem i regulatorer som ger direkta styrkommandon till ställdon i den fysiska infrastrukturen, se figur 1. Därmed kan felaktig kod i en inbyggd processor ge upphov till omedelbar fysisk skada, till exempel elavbrott. CERCES2 har inom detta område förfinat flera av de verktyg som utvecklades i CERCES.

Vi har utvecklat nya verktyg baserade på formell logik för att automatiskt, på matematiskt korrekt sätt, verifiera att program i inbyggda processorer har önskade egenskaper. Önskvärda egenskaper som vi verifierat är att en angripare inte kan få tillgång till en viss kritisk kryptografisk nyckel eller att tidsgränser som är viktiga för någon styrfunktion inte överskrids. Sådan verifikation är dock helt beroende av att modelleringen av den underliggande processorn stämmer med verkligheten. Tyvärr visar det sig ofta att så inte alls är fallet. De senaste åren, med framkomsten av ett stort antal hårdvarubaserade sårbarheter såsom Spectre, Meltdown och Foreshadow, har visat att vår kunskap om den faktiska processorhårdvaran är ofullständig, inte minst om så kallade caches. Dessa sårbarheter har visat sig vara högst allvarliga i många fall och går i allmänhet inte att blockera med enbart kodanalys. Konsekvensen är att säkerhetsanalysen, om den görs med enbart testning eller ens med formella metoder, blir felaktig. Under projektets gång har därför modelleringen och analysen av processorhårdvaran blivit en än mer viktig del av forskningsagendan, och resulterat i nya lovande modeller [12] och testmetoder [13].

En stor utmaning är komplexiteten av sökrymden eftersom modern hårdvara består av en mängd olika delfunktioner som kan interagera på en mängd olika sätt. Det är därför av vikt att sökprocessen kan ledas på ett intelligent sätt till att fokusera på en misstänkt del av sökrymden. I [14] presenterar vi en metod och verktyget SCAM-V som genom en stegvis och automatiserat förfining av sökinstrumentet presenterat i [13] lyckas isolera en ny variant av Spectre-sårbarheten. I relaterat arbete har vi studerat metoder för utveckling av säkra system med stegvis förfining. Ett fundamentalt och välkänt problem sedan 80-talet har varit att standardmetoder för stegvis förfining inte respekterar s.k. konfidentialitet. Därmed går det exempelvis att "korrekt" implementera ett nyckelhanteringsprotokoll som läcker alla sina nycklar, även om detta inte är tillåtet enligt specifikation. Detta är självfallet högst problematisk. I [15] föreslår vi en lösning på detta problem genom att kräva att förfiningssteg respekterar "okunskap" om givna data, till exempel en kryptografisk nyckel. Även om mer arbete med metoden behövs är detta ett viktigt framsteg, eftersom det tillåter erkända viktiga utvecklingsmetoder som stegvis förfining att appliceras även på säkra system.

I uppföljande arbete till [15] ger vi exempel på detta, genom att dels applicera metoden på ett enkelt nyckelhanteringsprotokoll, dels till att bevisa korrekthet och säkerhet för en femstegs "pipelined" processor implementerad i hårdvarudesign-språket *Verilog*.

En viktig metod för säkerhetsanalys på kodnivå är symbolisk exekvering. Med denna metod exekverar man program med vissa variabler ersatt av symboliska värden. Denna analysmetod möjliggör att enkelt analysera programberoenden, till exempel om ett visst värde som skickas på ett publikt nät är beroende av en annan, möjligen konfidentiell, variabel som till exempel en kryptografisk nyckel. Ett problem med metoden är att merparten sådana metoder underskattar mängden av möjliga exekveringar, vilket kan betyda att analysen blir behäftad med fel. Ett viktigt bidrag från CERCES2-projektet, och en kärna i doktorand Andreas Lindners arbete [16] är utvecklingen av en ny metod för symbolisk exekvering som bevisbart inte tillåter att exekveringar "glöms bort".

Fördjupad läsning

Innehållet i denna rapport baseras på de vetenskapliga artiklar som deltagarna i forskningsprojektet har skrivit och fått publicerat.

[1] ”Ökad säkerhet i industriella informations- och styrsystem”, MSB, 2021. (Länk: https://rib.msb.se/filer/pdf/29984.pdf).

[2] ”Cyber-Attack Against Ukrainian Critical Infrastructure”, CISA, US Department of Homeland Security, 2016. (Länk: https://us-cert.cisa.gov/ics/alerts/IR-ALERT-H-16-056-01).

[3] L. Salazar, *et al.*, “A Tale of Two Industroyers: It was the Season of Darkness”, IEEE Symposium on Security and Privacy (SP), 2024.

[4] H. Forssell and R. Thobaben, “Worst-case detection performance for distributed SIMO physical layer authentication,” IEEE Transactions on Communications, 2021.

[5] H. Forssell, R. Thobaben, “Worst-Case Detection Performance of Physical Layer Authentication Under Optimal MIMO Attacks”, IEEE International Conference on Communications (ICC), 2021.

[6] H. Forssell, R. Thobaben, “Delay performance of distributed physical layer authentication under Sybil attacks”, IEEE International Conference on Communications (ICC), 2021.

[7] S. Saritas, H. Forssell, R. Thobaben, “Adversarial attacks on CFO-based continuous physical layer authentication: A game theoretic study,” in Proc. IEEE International Conference on Communications (ICC), 2021.

[8] E. Shereen., K. Kazari, G. Dán “A Reinforcement Learning Approach to Undetectable Attacks against Automatic Generation Control”, IEEE Transactions on Smart Grid, 2024.

[9] S. Saritas, E. Shereen, H. Sandberg, G. Dán, “Adversarial Attacks on Continuous Authentication Security: A Dynamic Game Approach”, International Conference on Decision and Game Theory for Security (GameSec), 2019.

[10] D. Umsonst, H. Sandberg, “Experimental evaluation of sensor attacks and defense mechanisms in feedback systems”, Control Engineering Practice, 2022.

[11] D. Umsonst, S. Sarıtaş, G. Dán, and H. Sandberg, “A Bayesian Nash Equilibrium-Based Moving Target Defense against Stealthy Sensor Attacks”, IEEE Transactions on Automatic Control, 2024.

[12] R. Guanciale, M. Balliu, M. Dam, “InSpectre: Breaking and Fixing Microarchitectural Vulnerabilities by Formal Analysis”, 2020 ACM Conference on Computer and Communications Security, 2020.

[13] H. Nemati, P. Buiras, A. Lindner, R. Guanciale, S. Jacobs, “Validation of Abstract Side-Channel Models for Computer Architectures”, 32nd International Conference on Computer Aided Verification, CAV, 2020.

[14] P. Buiras, H. Nemati, A. Lindner, R. Guanciale, "Validation of side-channel models via observational refinement", 54th IEEE/ACM International Symposium on Microarchitecture, MICRO'21, 2021.
[15] C. Baumann, M. Dam, R. Guanciale, H. Nemati, "On compositional information flow aware refinement", IEEE Computer Security Foundations Symposium, 2021.
[16] A. Lindner: Proving safety and security of binary programs. PhD thesis, KTH Royal Institute of Technology, 2023. (Länk: https://www.diva-portal.org/smash/get/diva2:1755837/FULLTEXT01.pdf).

Myndigheten för samhällsskydd och beredskap

I samarbete med:

651 81 Karlstad Tel 0771-240 240 **www.msb.se**
Publikationsnummer MSB2505 – december 2024