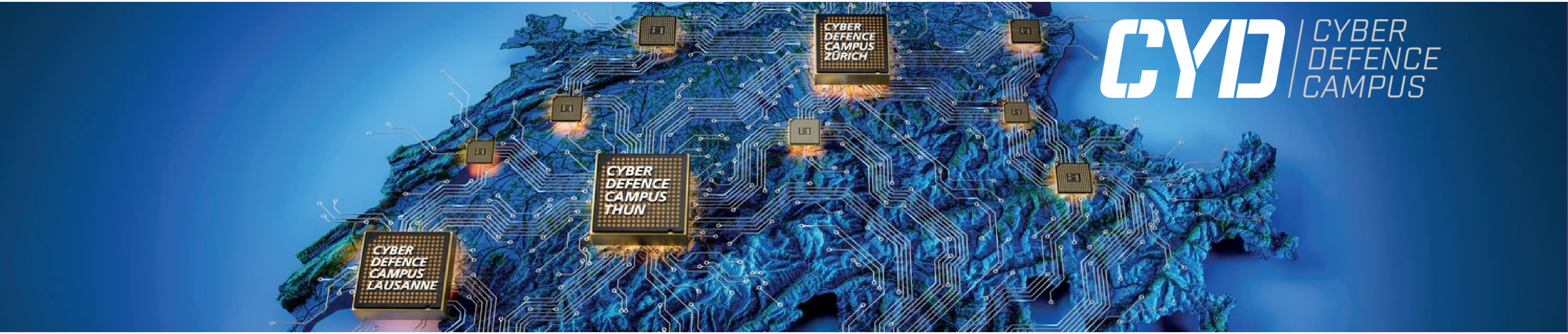
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für Verteidigung,
Bevölkerungsschutz und Sport VBS
Bundesamt für Rüstung - armasuisse ar

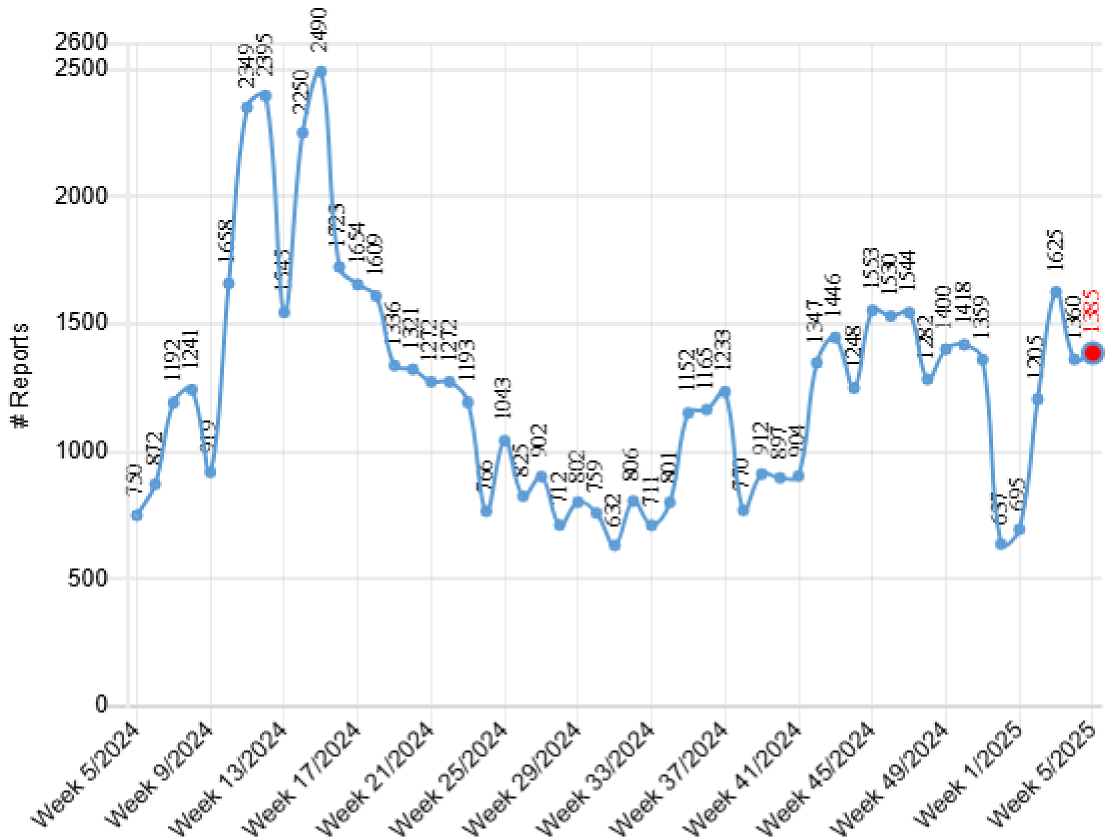

CYD CYBER DEFENCE CAMPUS
CYBER DEFENCE CAMPUS ZÜRICH
CYBER DEFENCE CAMPUS THUN
CYBER DEFENCE CAMPUS LAUSANNE

Towards a fully automated Blue Team at Locked Shields

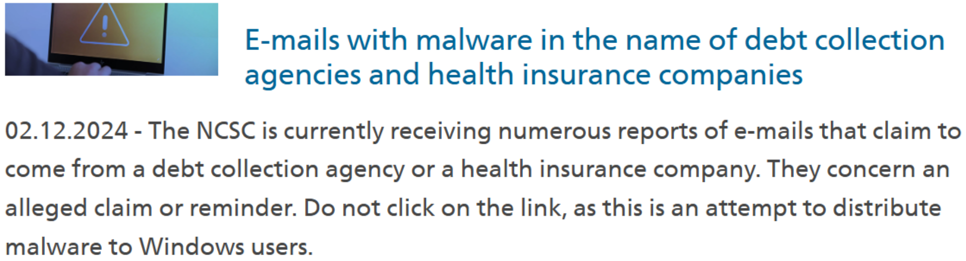
Roland Meier
AMLD 2025, February 13, 2025

Every week, the NCSC receives >1`000 reports about cyber incidents

Chart 1 - NCSC.ch: reports received

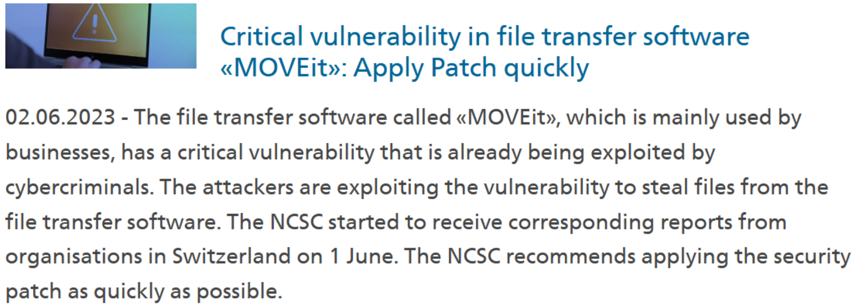


Every week, the NCSC receives >1`000 reports about cyber incidents



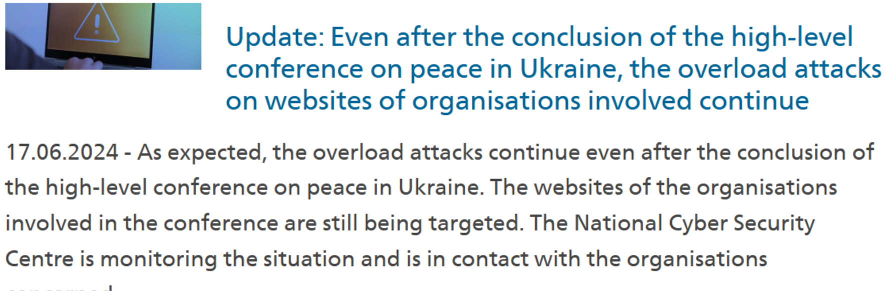
E-mails with malware in the name of debt collection agencies and health insurance companies

02.12.2024 - The NCSC is currently receiving numerous reports of e-mails that claim to come from a debt collection agency or a health insurance company. They concern an alleged claim or reminder. Do not click on the link, as this is an attempt to distribute malware to Windows users.



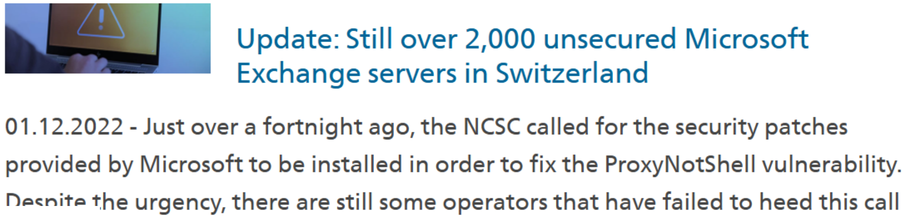
Critical vulnerability in file transfer software «MOVEit»: Apply Patch quickly

02.06.2023 - The file transfer software called «MOVEit», which is mainly used by businesses, has a critical vulnerability that is already being exploited by cybercriminals. The attackers are exploiting the vulnerability to steal files from the file transfer software. The NCSC started to receive corresponding reports from organisations in Switzerland on 1 June. The NCSC recommends applying the security patch as quickly as possible.



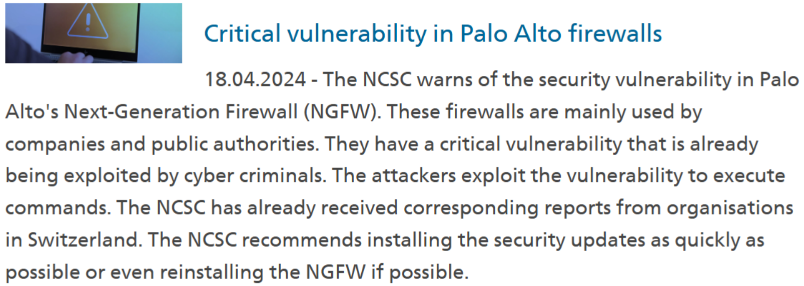
Update: Even after the conclusion of the high-level conference on peace in Ukraine, the overload attacks on websites of organisations involved continue

17.06.2024 - As expected, the overload attacks continue even after the conclusion of the high-level conference on peace in Ukraine. The websites of the organisations involved in the conference are still being targeted. The National Cyber Security Centre is monitoring the situation and is in contact with the organisations concerned.



Update: Still over 2,000 unsecured Microsoft Exchange servers in Switzerland

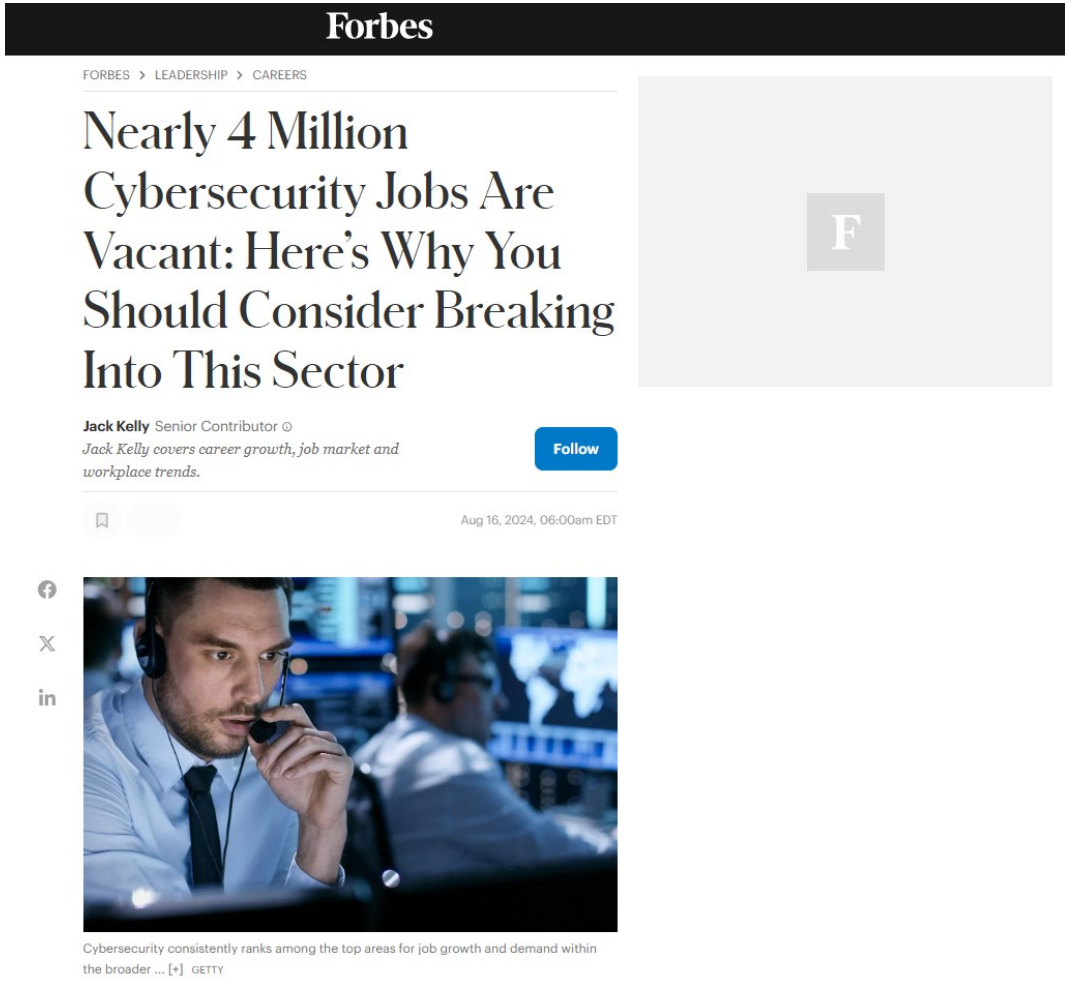
01.12.2022 - Just over a fortnight ago, the NCSC called for the security patches provided by Microsoft to be installed in order to fix the ProxyNotShell vulnerability. Despite the urgency, there are still some operators that have failed to heed this call Therefore, the NCSC has sent more than 2,000 registered letters to those ed, urging them to act now.



Critical vulnerability in Palo Alto firewalls

18.04.2024 - The NCSC warns of the security vulnerability in Palo Alto's Next-Generation Firewall (NGFW). These firewalls are mainly used by companies and public authorities. They have a critical vulnerability that is already being exploited by cyber criminals. The attackers exploit the vulnerability to execute commands. The NCSC has already received corresponding reports from organisations in Switzerland. The NCSC recommends installing the security updates as quickly as possible or even reinstalling the NGFW if possible.

There are not enough experts in this field



Forbes

FORBES › LEADERSHIP › CAREERS

Nearly 4 Million Cybersecurity Jobs Are Vacant: Here's Why You Should Consider Breaking Into This Sector

Jack Kelly Senior Contributor

Jack Kelly covers career growth, job market and workplace trends.

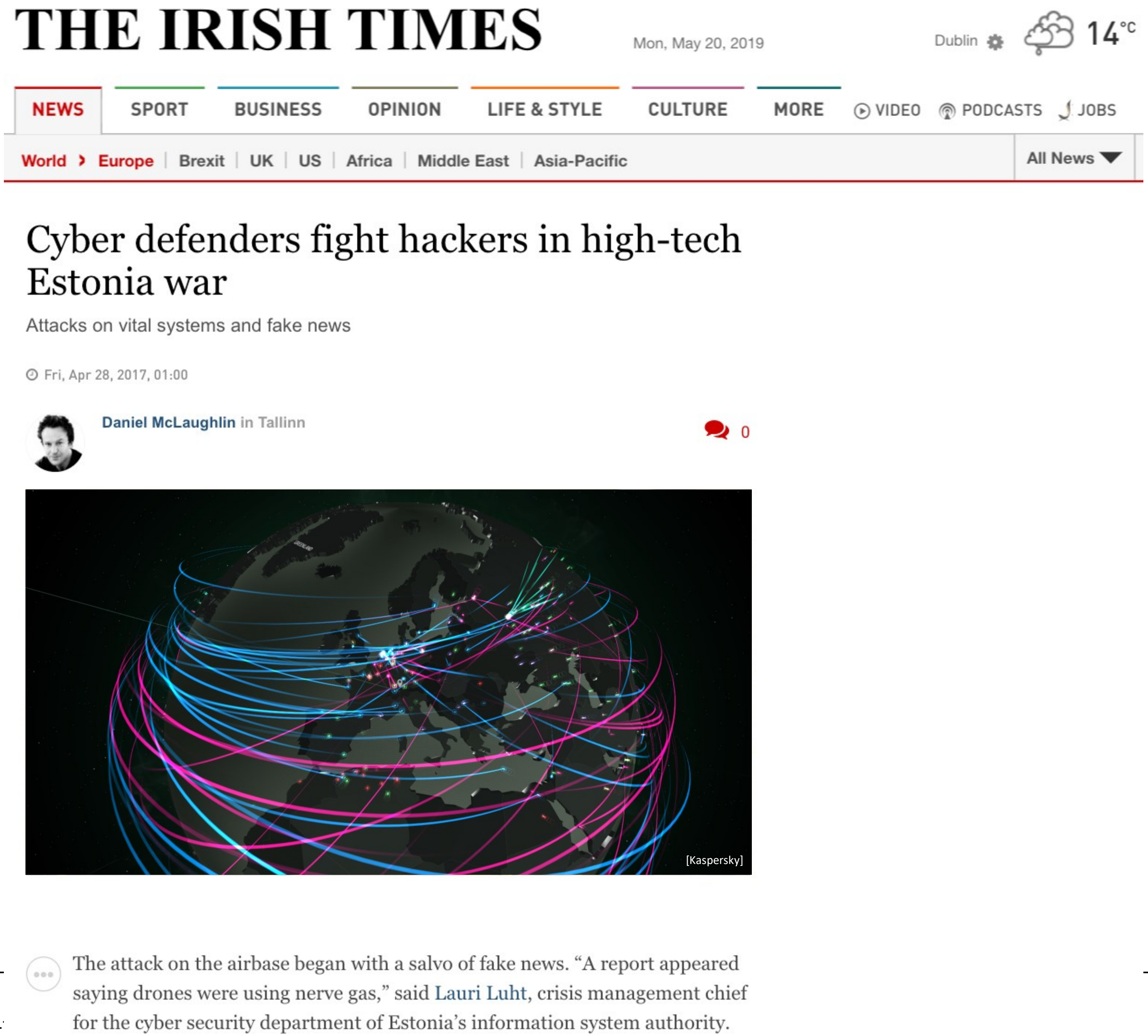
Aug 16, 2024, 06:00am EDT

Cybersecurity consistently ranks among the top areas for job growth and demand within the broader ... [+] GETTY

AI is everywhere…

GENIUS X
THE REVOLUTIONARY GENIUS X WITH ARTIFICIAL INTELLIGENCE
Recognizes your brushing style. Guides you to brush better every day.
SHOP NOW
Oral-B
Oral-B
Oral-B

… why not in cyber defense?



Cyber defenders fight hackers in high-tech Estonia war

Attacks on vital systems and fake news

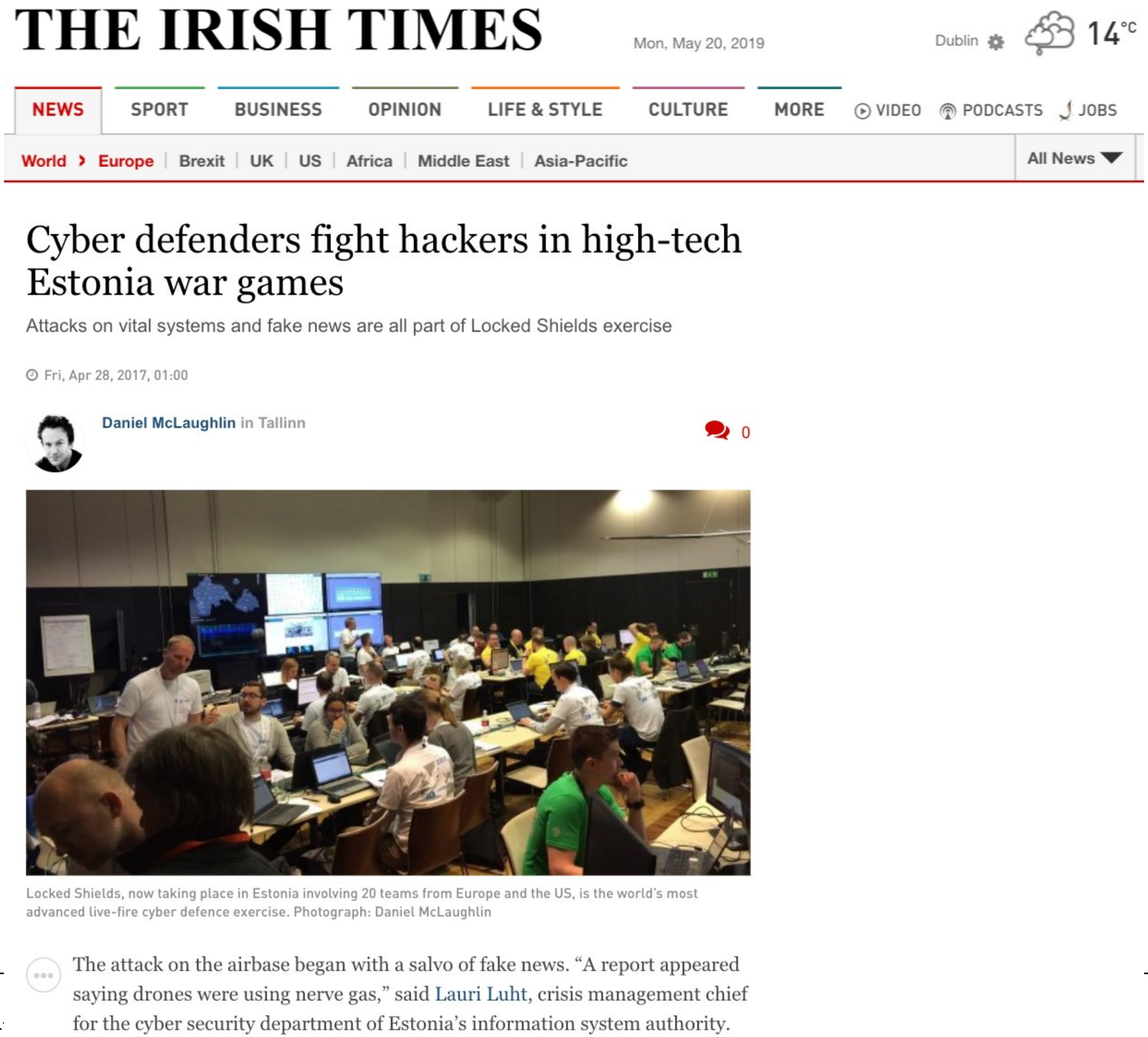
Fri, Apr 28, 2017, 01:00

Daniel McLaughlin in Tallinn

0

[Kaspersky]

The attack on the airbase began with a salvo of fake news. “A report appeared saying drones were using nerve gas,” said Lauri Luht, crisis management chief for the cyber security department of Estonia’s information system authority.



Cyber defenders fight hackers in high-tech Estonia war games

Attacks on vital systems and fake news are all part of Locked Shields exercise

Fri, Apr 28, 2017, 01:00

Daniel McLaughlin in Tallinn

Locked Shields, now taking place in Estonia involving 20 teams from Europe and the US, is the world's most advanced live-fire cyber defence exercise. Photograph: Daniel McLaughlin

The attack on the airbase began with a salvo of fake news. "A report appeared saying drones were using nerve gas," said Lauri Luht, crisis management chief for the cyber security department of Estonia's information system authority.

Locked Shields is the largest live-fire global cyber defense exercise

Picture: NATO CCDCOE

Locked Shields is the largest live-fire global cyber defense exercise

- Red Team vs. Blue Team exercise

Attackers
1 Team

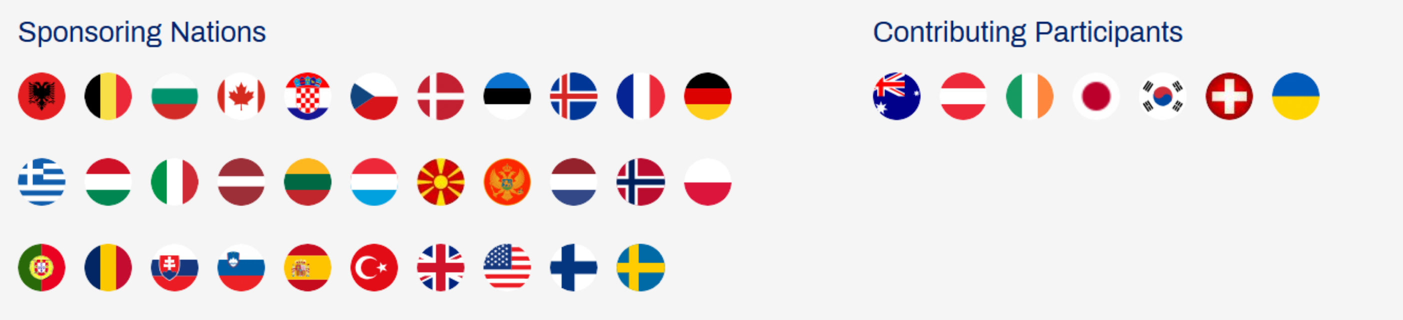
Defenders
1 Team / country

- > 1’000 experts from 30 nations
- > 4’000 systems
- > 2’500 attacks

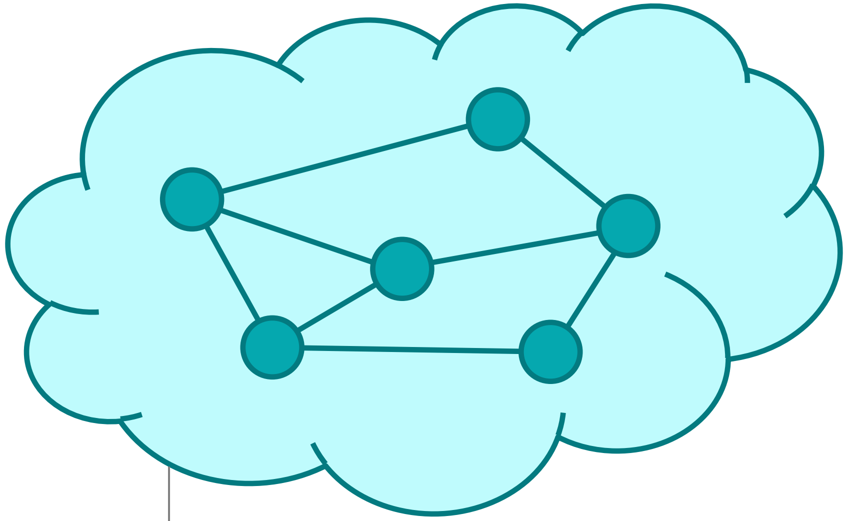
Picture: NATO CCDCOE

Locked Shields is organized by the NATO CCDCoE

KEYSTONE | PETER KLAUNZER

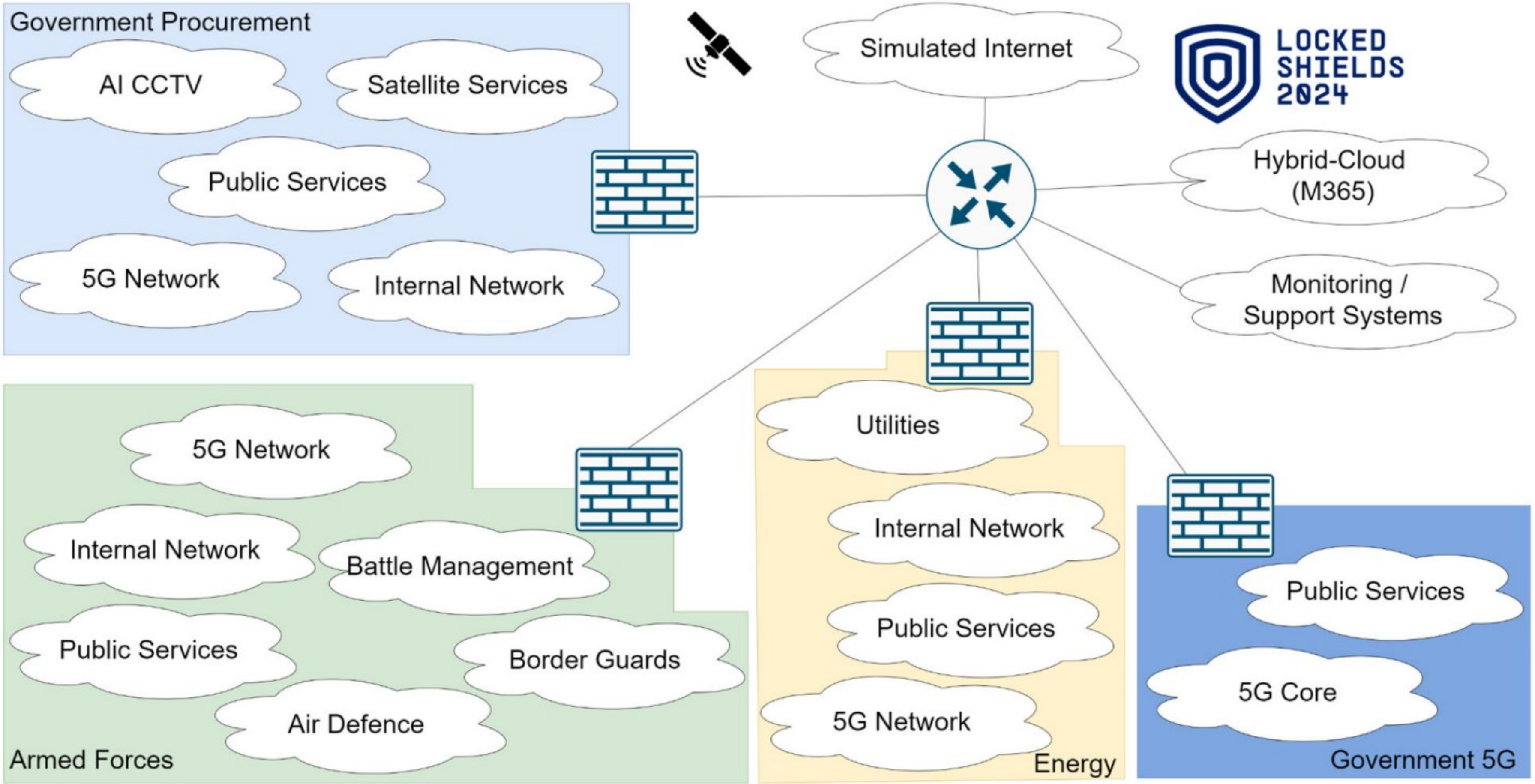


Each Blue Team is responsible for its network ("Gamenet")

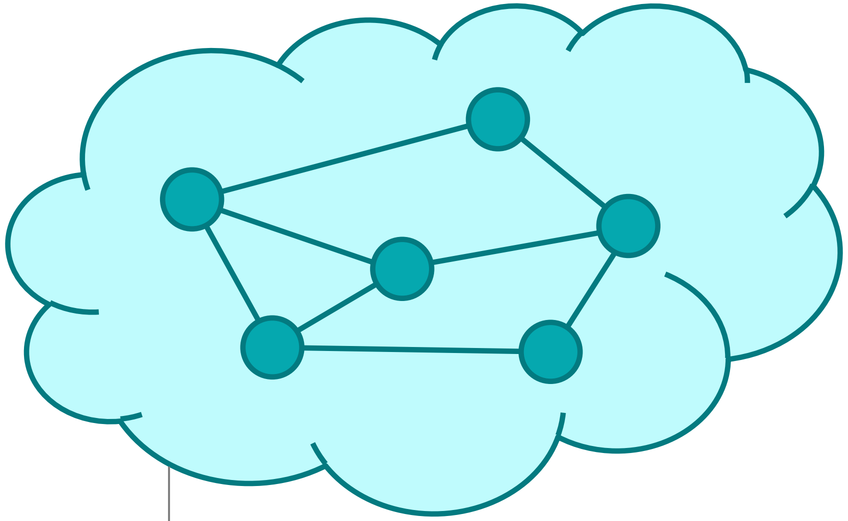


e.g., a military air base

The Gamenet consists of a large variety of systems



Each Blue Team is responsible for its network (“Gamenet”)



e.g., a military air base

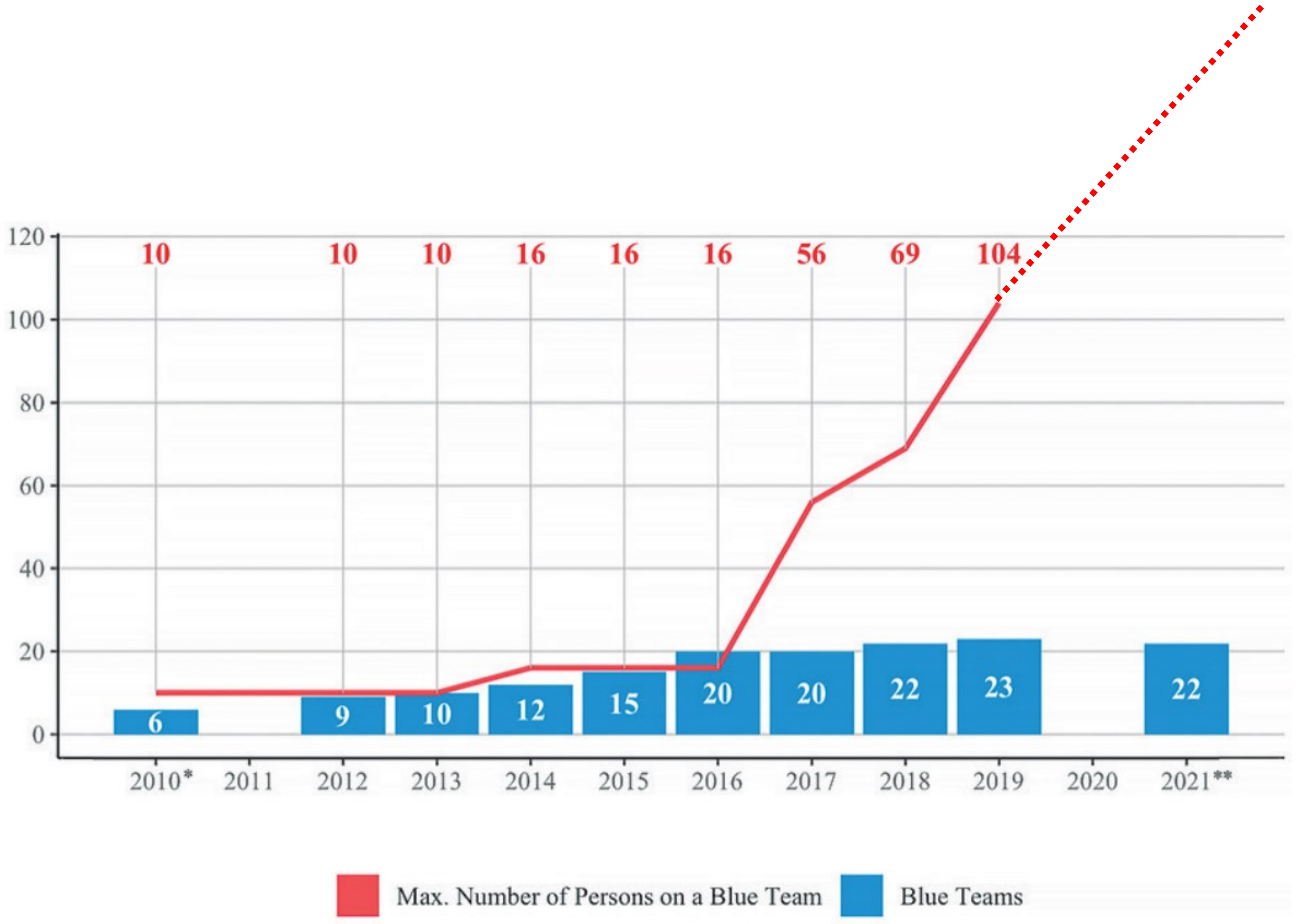
3 main tasks:

- Perform initial hardening
- Defend against attacks
- Communicate with other teams

Besides defending its network, a Blue Team needs to communicate with other teams

User Simulation Team	Read, address and respond to support tickets
Yellow Team	Provide periodic reports
White team	Voice or video calls through the Gamenet
Green Team	Gamenet status, reverting of devices

The number of people required in a Blue Team continuously increases



[Smeets, Max. "The role of military cyber exercises: A case study of Locked Shields." *CyCon 2022*]

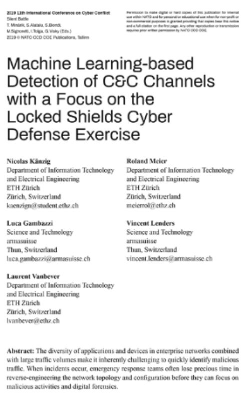
Research goals

How can automation / AI help for cyber defense?

And eventually…

What would it take to have a fully automated Blue Team in a future iteration of Locked Shields?

The history



2019 11th International Conference on Cyber Conflict
Silent Battle
T. Minárik, S. Alatalu, S. Biondi,
M. Signoretti, I. Tolga, G. Visky (Eds.)
2019 © NATO CCD COE Publications, Tallinn

Machine Learning-based Detection of C&C Channels with a Focus on the Locked Shields Cyber Defense Exercise

Nicolas Känzig
Department of Information Technology and Electrical Engineering
ETH Zürich
Zürich, Switzerland
kaenzign@student.ethz.ch

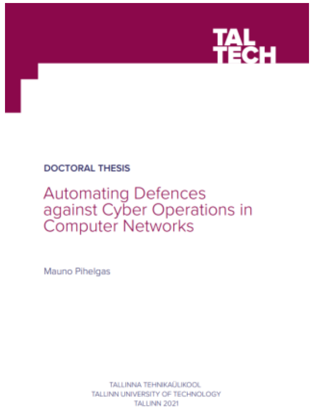
Roland Meier
Department of Information Technology and Electrical Engineering
ETH Zürich
Zürich, Switzerland
meierrol@ethz.ch

Luca Gambazzi
Science and Technology
armasuisse
Thun, Switzerland
luca.gambazzi@armasuisse.ch

Vincent Lenders
Science and Technology
armasuisse
Thun, Switzerland
vincent.lenders@armasuisse.ch

Laurent Vanbever
Department of Information Technology and Electrical Engineering
ETH Zürich
Zürich, Switzerland
lvanbever@ethz.ch

Abstract: The diversity of applications and devices in enterprise networks combined with large traffic volumes make it inherently challenging to quickly identify malicious traffic. When incidents occur, emergency response teams often lose precious time in reverse-engineering the network topology and configuration before they can focus on malicious activities and digital forensics.



TAL TECH

DOCTORAL THESIS

Automating Defences against Cyber Operations in Computer Networks

Mauno Pihelgas

TALLINNA TEHNIKAÜLIKOOL
TALLINN UNIVERSITY OF TECHNOLOGY
TALLINN 2021

Towards an Active, Autonomous and Intelligent Cyber Defense of Military Systems: the NATO AICA Reference Architecture

Paul Théron, Alexander Kott, Martin Drašar, Krzysztof Rzadca, Benoît LeBlanc, Mauno Pihelgas, Luigi Mancini, Agostino Panico

2021 13th International Conference on Cyber Conflict
Going Viral
T. Jančárková, L. Lindström, M. Visky, P. Zotz (Eds.)
2021 © NATO CCDCOE Publications, Tallinn

Towards an AI-powered Player in Cyber Defence Exercises

Roland Meier
Department of Information Technology and Electrical Engineering
ETH Zürich
Zürich, Switzerland
meierrol@ethz.ch

Artūrs Lavrenovs
NATO CCDCOE
Tallinn, Estonia
arturs.lavrenovs@ccdcoe.org

Kimmo Heinäaro
NATO CCDCOE
Tallinn, Estonia
kimmo.heinaaro@mil.fi

Luca Gambazzi
Science and Technology
armasuisse
Thun, Switzerland
luca.gambazzi@armasuisse.ch

Vincent Lenders
Science and Technology
armasuisse
Thun, Switzerland
vincent.lenders@armasuisse.ch

Abstract: Cyber attacks are becoming increasingly frequent, sophisticated, and stealthy. This makes it harder for cyber defence teams to keep up, forcing them to automate their defence capabilities in order to improve their reactivity and efficiency. Therefore, we propose a fully automated cyber defence framework that no longer needs support from humans to detect and mitigate attacks within a complex infrastructure. We design our framework based on a real-world case – Locked Shields – the world's largest cyber defence exercise. In this exercise, teams have to defend their networked infrastructure against attacks, while maintaining operational services for their users. Our framework architecture connects various cyber sensors with network, device, application, and user actuators through an artificial intelligence (AI)-powered automated team in order to dynamically secure the cyber environment. To the best of our knowledge, our framework is the first attempt towards a fully automated cyber defence team that aims at protecting complex environments from sophisticated attacks.

Keywords: *artificial intelligence, automation, Locked Shields, cyber defence, security*

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Towards Generalizing Machine Learning Models to Detect Command and Control Attack Traffic

Lina Gehri
ETH Zurich
Department of Electrical Engineering and Information Technology
Zurich, Switzerland
lina.gehri@gmail.com

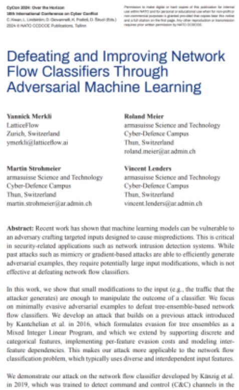
Roland Meier
Cyber-Defence Campus
armasuisse Science and Technology
Thun, Switzerland
roland.meier@ar.admin.ch

Daniel Hulliger
Cyber-Defence Campus
armasuisse Science and Technology
Thun, Switzerland
daniel.hulliger@ar.admin.ch

Vincent Lenders
Cyber-Defence Campus
armasuisse Science and Technology
Thun, Switzerland
vincent.lenders@ar.admin.ch

Abstract: Identifying compromised hosts from network traffic traces has become challenging because benign and malicious traffic is encrypted, and both use the same protocols and ports. Machine learning-based anomaly detection models have been proposed to address this challenge by classifying malicious traffic based on network flow features learned from historical patterns. Previous work has shown that such models successfully identify compromised hosts in the same network environment in which they were trained. However, cyber incidence response teams often have to look for intrusions in foreign networks, and we have found that learned models often fail to generalize to different network conditions. In this paper, we analyse the root cause of this problem using five network traces collected from different years and teams of Locked Shields, the world's largest live-fire cyber defence exercise. We then explore techniques to make machine learning models generalize better to unknown network environments and evaluate their accuracy.

Keywords: *machine learning, traffic classification, network security, command and control, Locked Shields*



CyCon 2024: Over the Horizon
16th International Conference on Cyber Conflict
C. Kwan, L. Lindström, D. Giovannelli, K. Podiņš, D. Štrucl (Eds.)
2024 © NATO CCDCOE Publications, Tallinn

Defeating and Improving Network Flow Classifiers Through Adversarial Machine Learning

Yannick Merkli
LatticeFlow
Zurich, Switzerland
ymerkli@latticeflow.ai

Roland Meier
armasuisse Science and Technology
Cyber-Defence Campus
Thun, Switzerland
roland.meier@ar.admin.ch

Martin Strohmeier
armasuisse Science and Technology
Cyber-Defence Campus
Thun, Switzerland
martin.strohmeier@ar.admin.ch

Vincent Lenders
armasuisse Science and Technology
Cyber-Defence Campus
Thun, Switzerland
vincent.lenders@ar.admin.ch

Abstract: Recent work has shown that machine learning models can be vulnerable to an adversary crafting targeted inputs designed to cause mispredictions. This is critical in security-related applications such as network intrusion detection systems. While past attacks such as mimicry or gradient-based attacks are able to efficiently generate adversarial examples, they require potentially large input modifications, which is not effective at defeating network flow classifiers.

In this work, we show that small modifications to the input (e.g., the traffic that the attacker generates) are enough to manipulate the outcome of a classifier. We focus on minimally evasive adversarial examples to defeat tree-ensemble-based network flow classifiers. We develop an attack that builds on a previous attack introduced by Kantchelian et al. in 2016, which formulates evasion for tree ensembles as a Mixed Integer Linear Program, and which we extend by supporting discrete and categorical features, implementing per-feature evasion costs and modeling inter-feature dependencies. This makes our attack more applicable to the network flow classification problem, which typically uses diverse and interdependent input features.

We demonstrate our attack on the network flow classifier developed by Känzig et al. in 2019, which was trained to detect command and control (C&C) channels in the

Contents lists available at ScienceDirect

Journal of Information Security and Applications

journal homepage: www.elsevier.com/locate/jisa

ELSEVIER

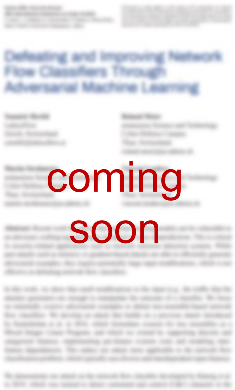
LSPR23: A novel IDS dataset from the largest live-fire cybersecurity exercise

Allard Dijk, Emre Halisdemir, Cosimo Melella, Alari Schu, Mauno Pihelgas, Roland Meier

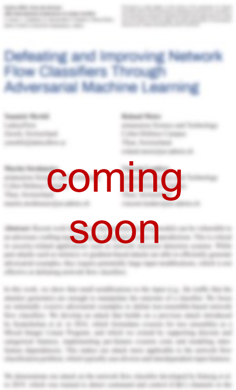
ARTICLE INFO

ABSTRACT

1. Introduction

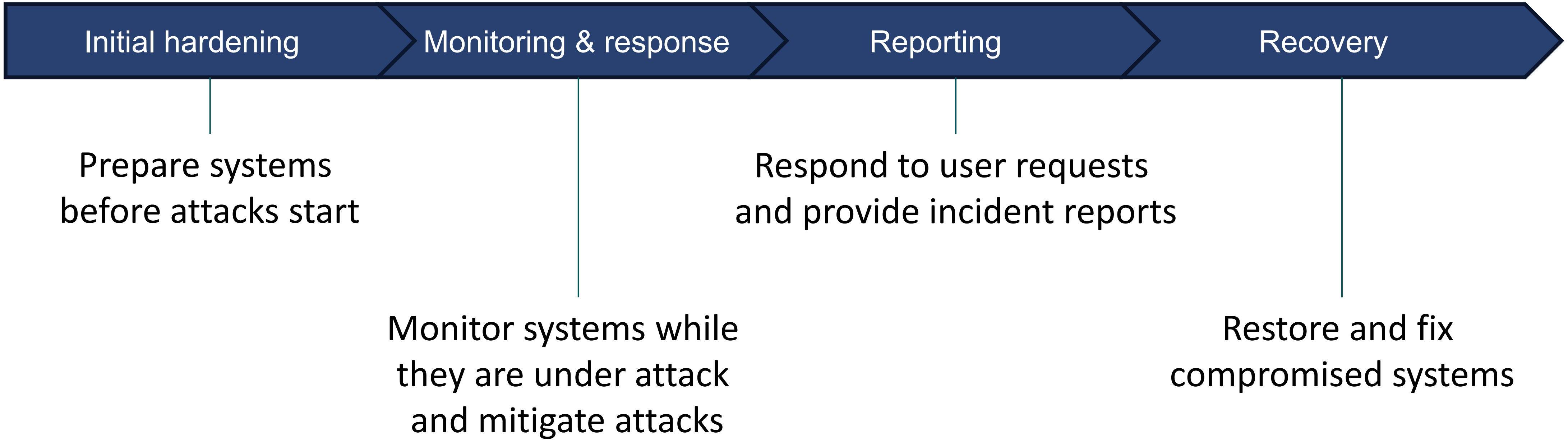


coming soon



coming soon

Four stages in the exercise



Initial hardening

Monitoring & response

Reporting

Recovery

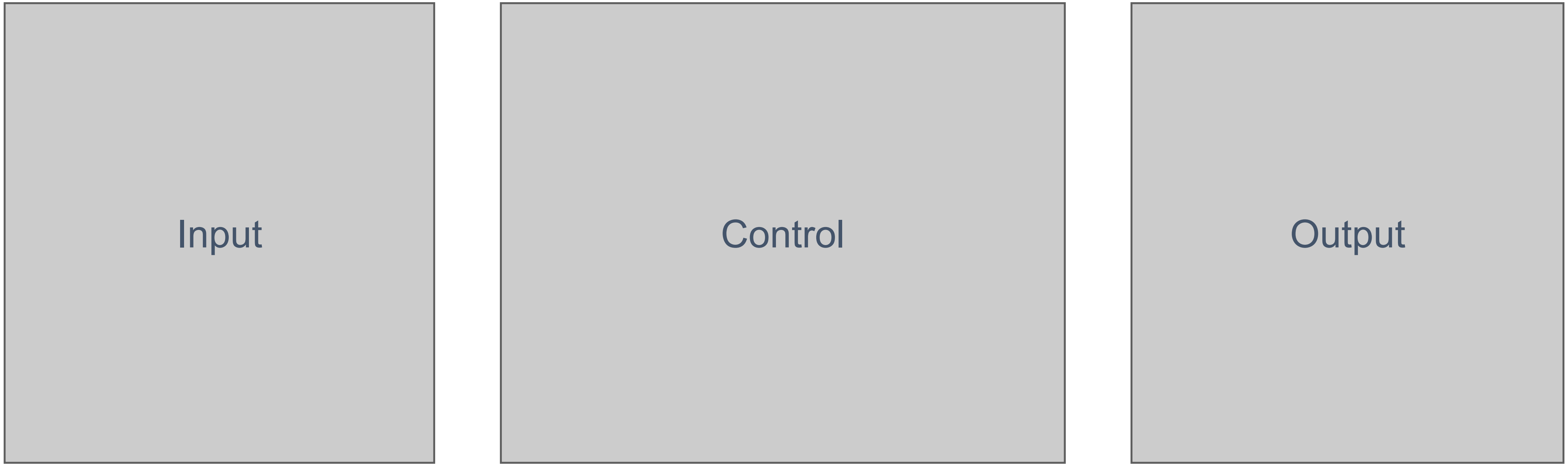
Prepare systems before attacks start

Monitor systems while they are under attack and mitigate attacks

Respond to user requests and provide incident reports

Restore and fix compromised systems

Automated Blue Team framework overview

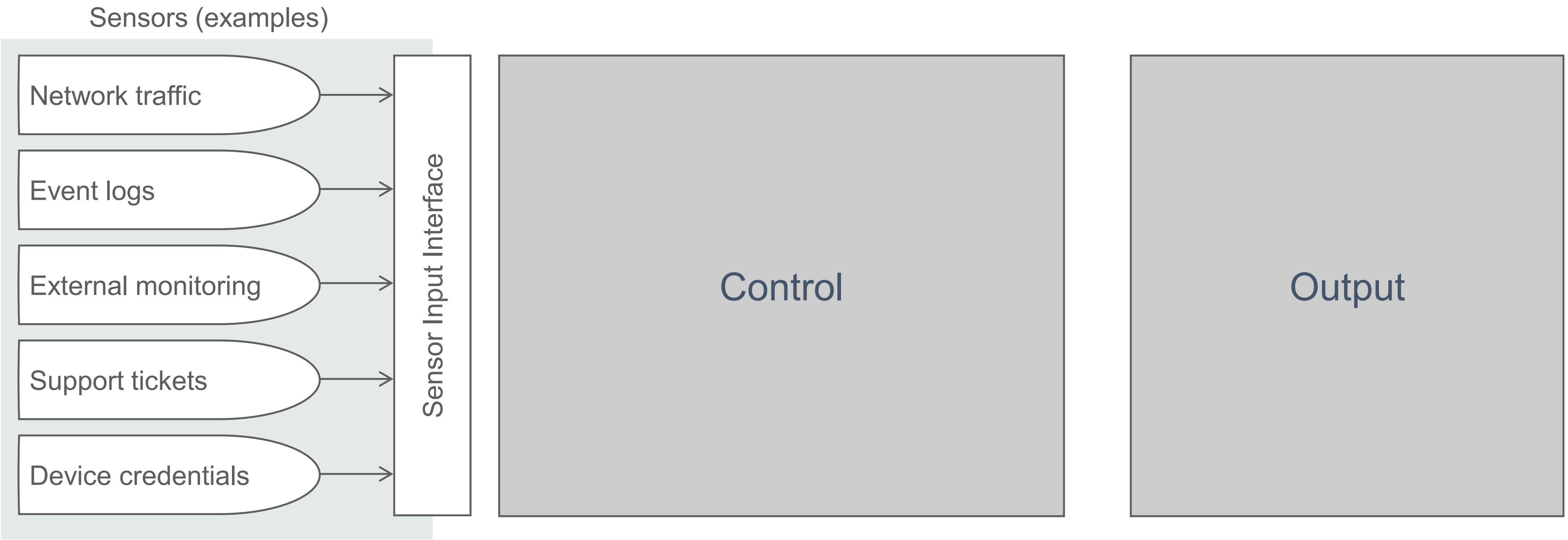


Input

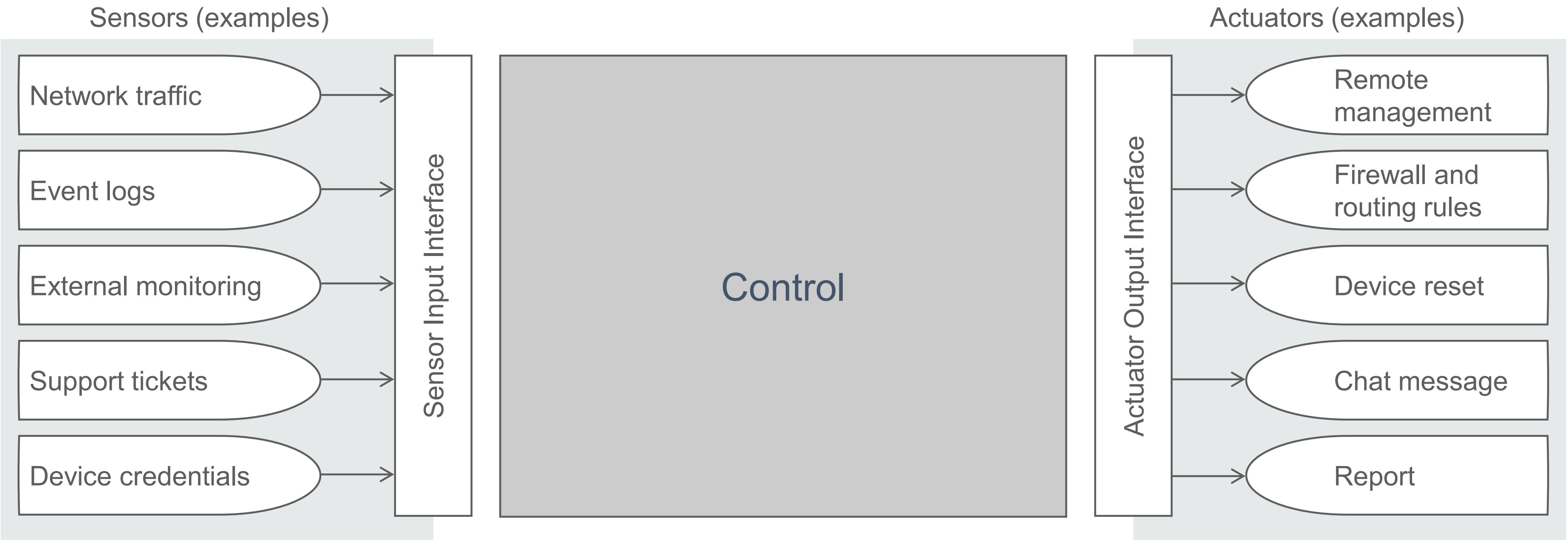
Control

Output

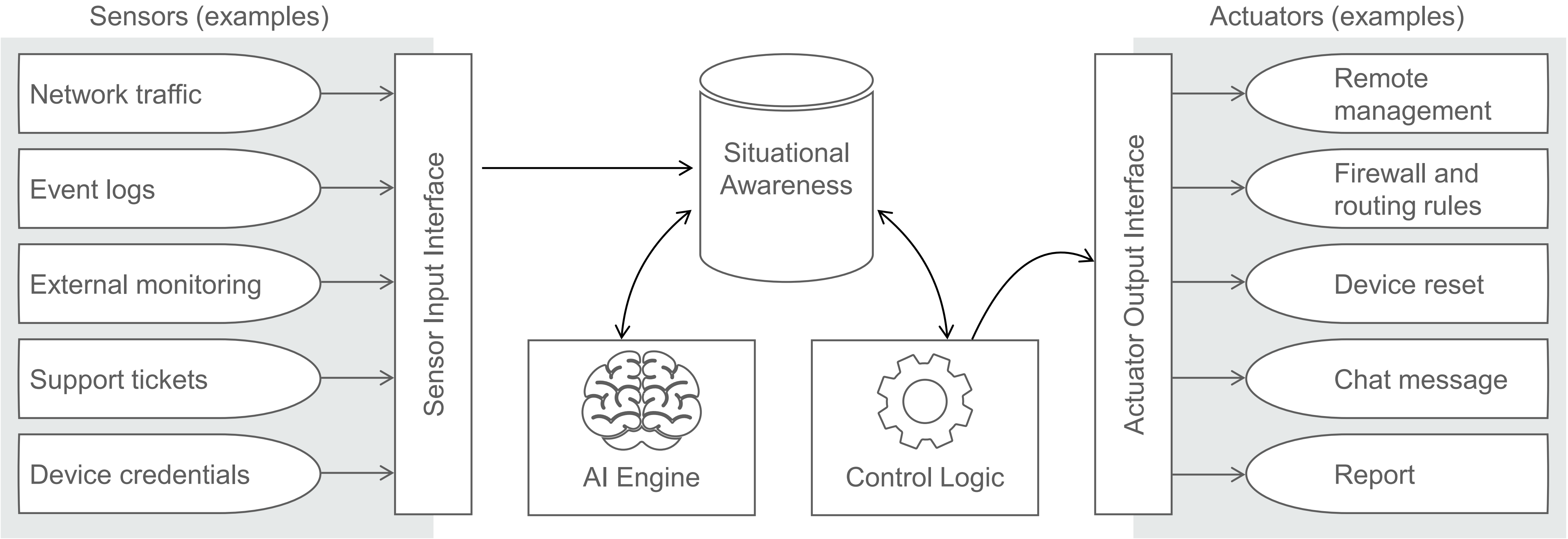
Automated Blue Team framework overview



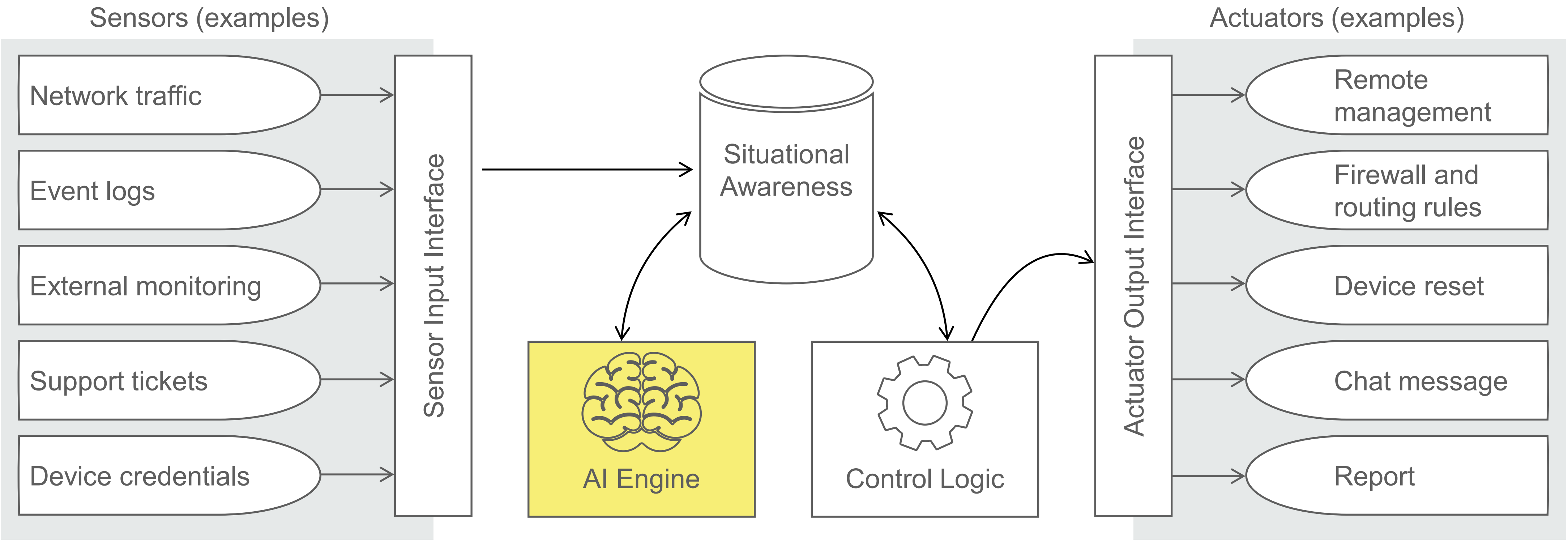
Automated Blue Team framework overview



Automated Blue Team framework overview



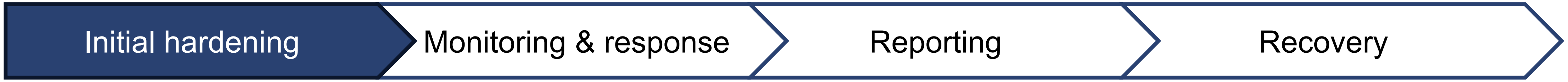
Automated Blue Team framework overview



Five tasks for AI

Identification / classification	What is it?
Categorisation	What belongs together?
Assessment	What is important?
Recommendation	What should be done?
Prediction	What will happen?

AI for initial hardening



Initial hardening

Monitoring & response

Reporting

Recovery

Identification / classification

Categorisation

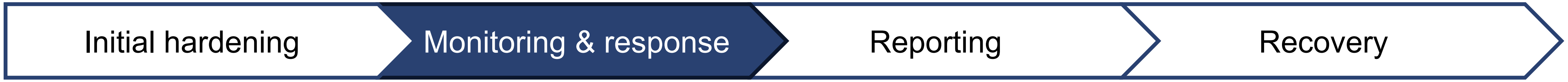
Find groups of similar devices

Assessment

Recommendation

Prediction

AI for monitoring and response



Identification / classification

Categorisation

Assessment

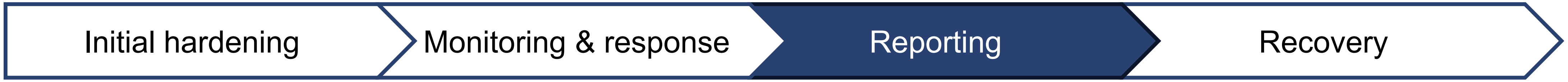
Recommendation

Prediction

Detect malicious network traffic

Detect malicious patterns in log files

AI for reporting



Initial hardening

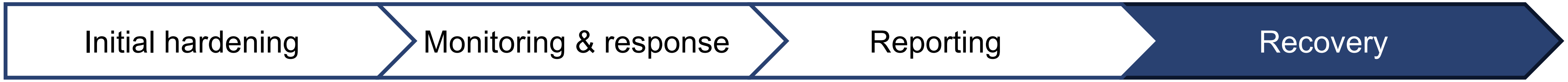
Monitoring & response

Reporting

Recovery

Identification / classification	
Categorisation	Link support tickets and monitoring alerts
Assessment	Prioritise support tickets
Recommendation	Formulate response to support tickets
Prediction	Predict impact on the scoring

AI for recovery



Initial hardening

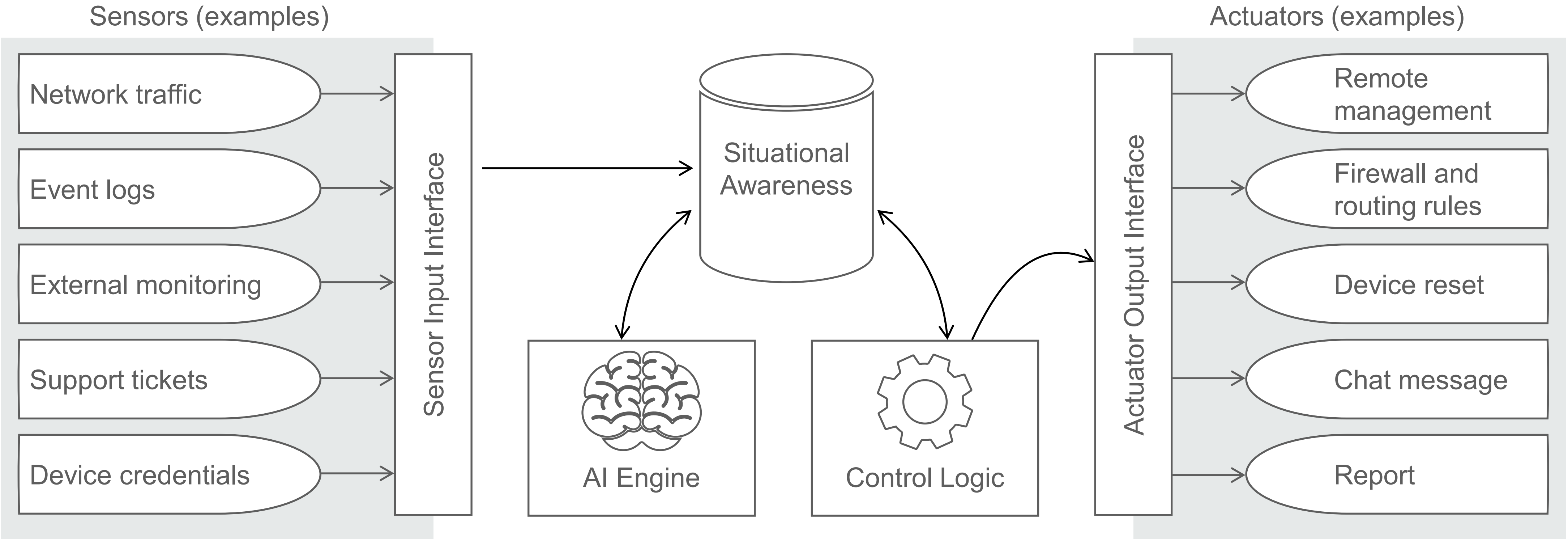
Monitoring & response

Reporting

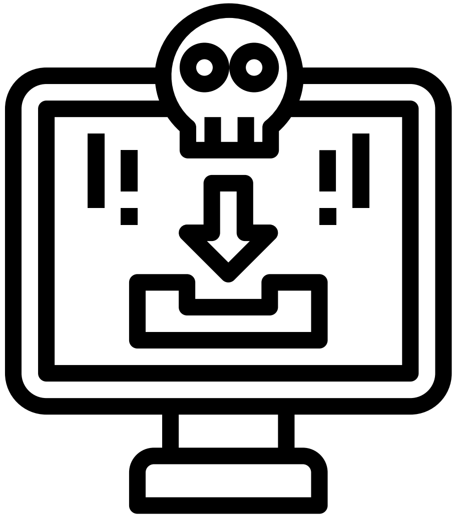
Recovery

Identification / classification	Find devices that need to be recovered
Categorisation	Find similar devices as a template for recovery
Assessment	
Recommendation	
Prediction	

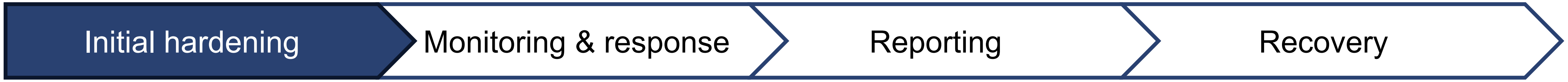
Automated Blue Team framework overview



Case study: A user mistakenly downloads and executes a malicious file



Case study



Initial hardening

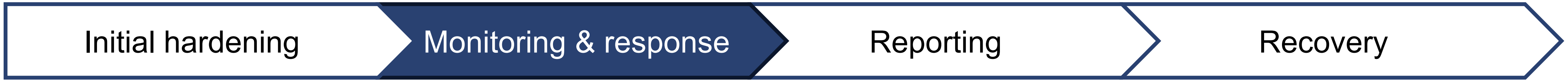
Monitoring & response

Reporting

Recovery

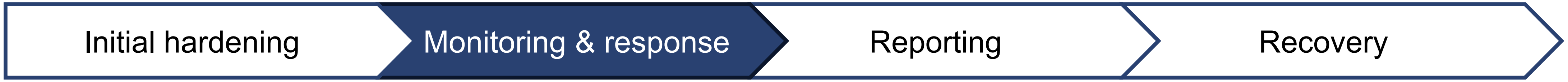
- Configure all clients to send HTTP(s) traffic via a proxy
- Enable detailed logging and send logs to a central server
- Set up recording of all network traffic and feature extraction

Case study



- Proxy detects the malicious payload
- Logging reports the execution of an unknown file
- Sniffer detects connection to C&C server

Case study


Initial hardening

Monitoring & response

Reporting

Recovery

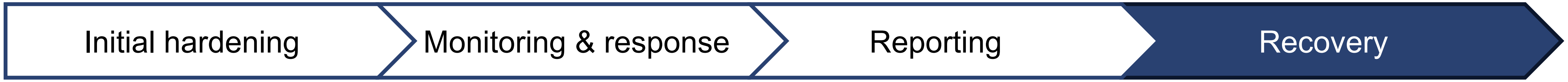
- Proxy detects the malicious payload → Remove payload
- Logging reports the execution of an unknown file → Block execution
- Sniffer detects connection to C&C server → Drop packets

Case study

Initial hardening → Monitoring & response → **Reporting** → Recovery

- Human-readable report with information about the incident
  - Malware source (compromised webserver)
  - Malware type
  - …

Case study



Initial hardening

Monitoring & response

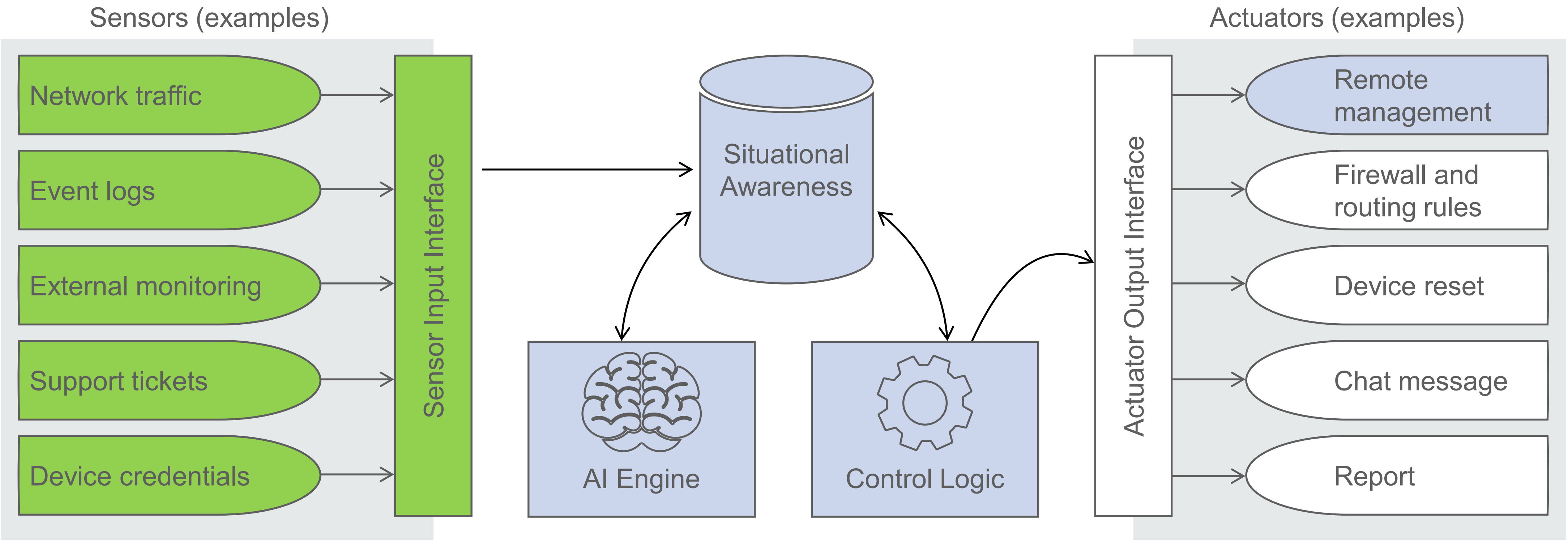
Reporting

Recovery

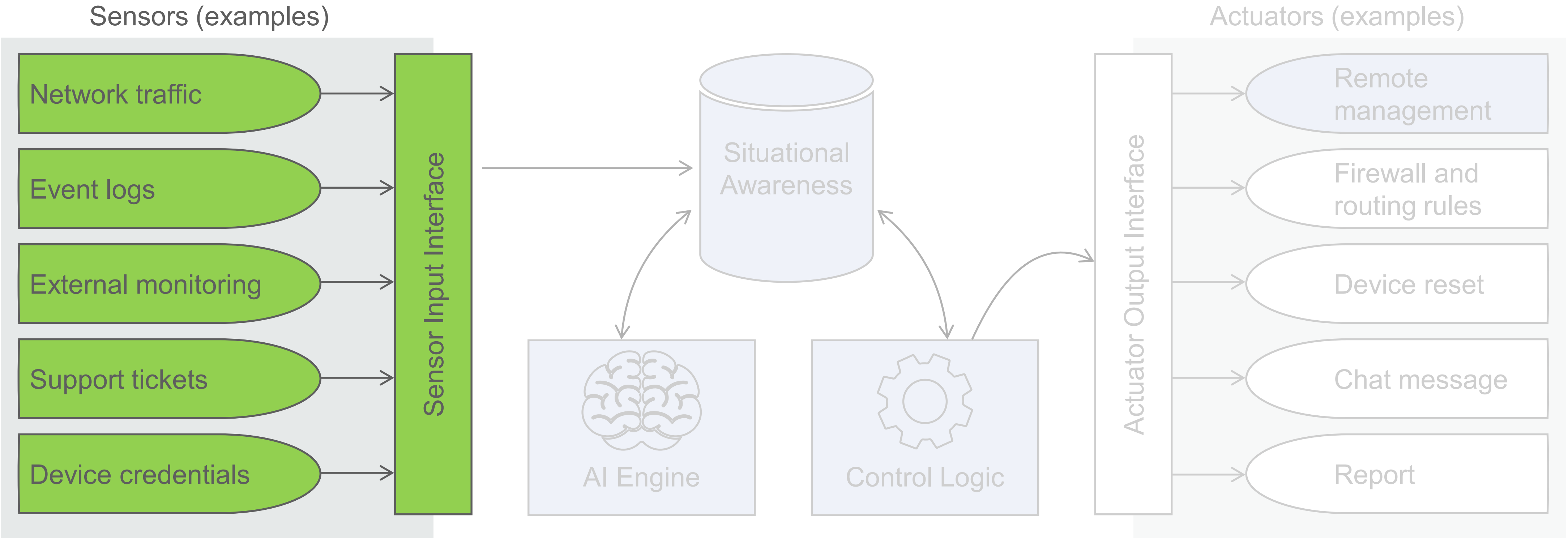
- Restore a device if the malware was executed

  E.g., from a backup

Current status of the project

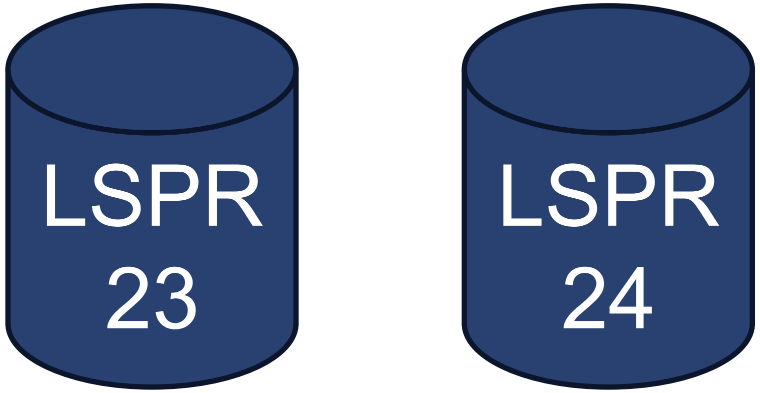


Current status of the project



Training and testing AI models requires high-quality datasets

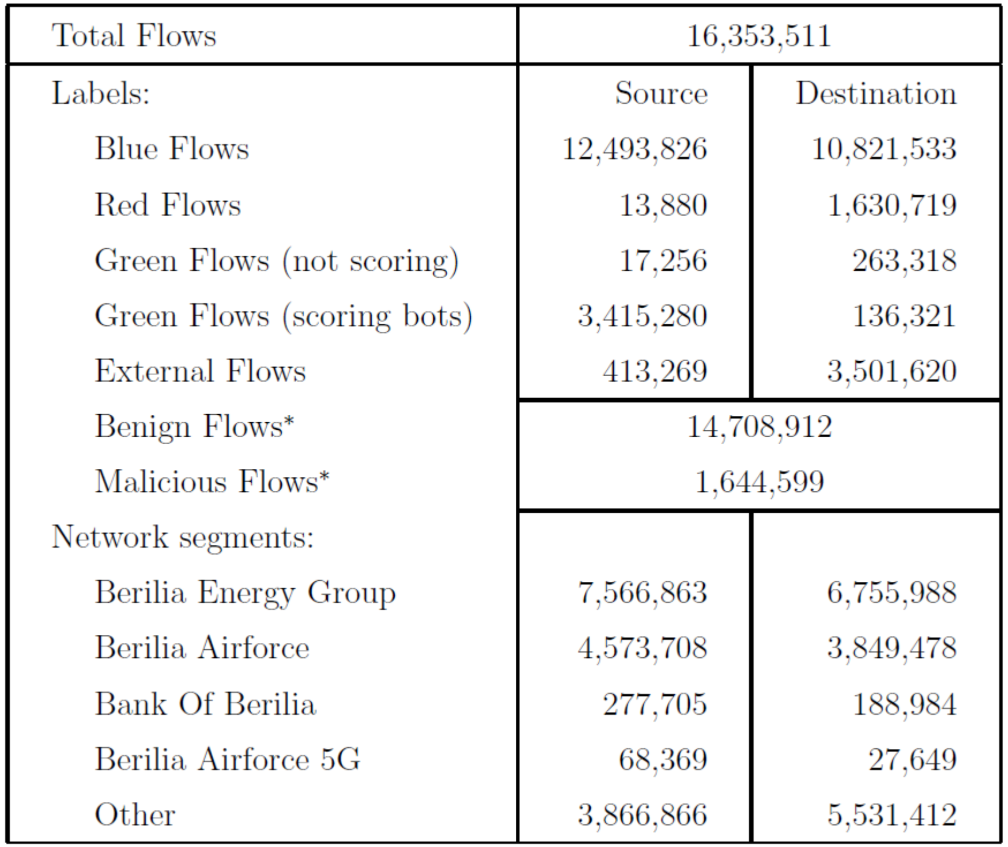
- We work with data from the Swiss Blue Team and other collaborating nations
- Since 2023, we (researchers) participate as a separate Blue Team in Locked Shields for data collection
- Datasets collected in 2023 and 2024 are (soon) publicly available



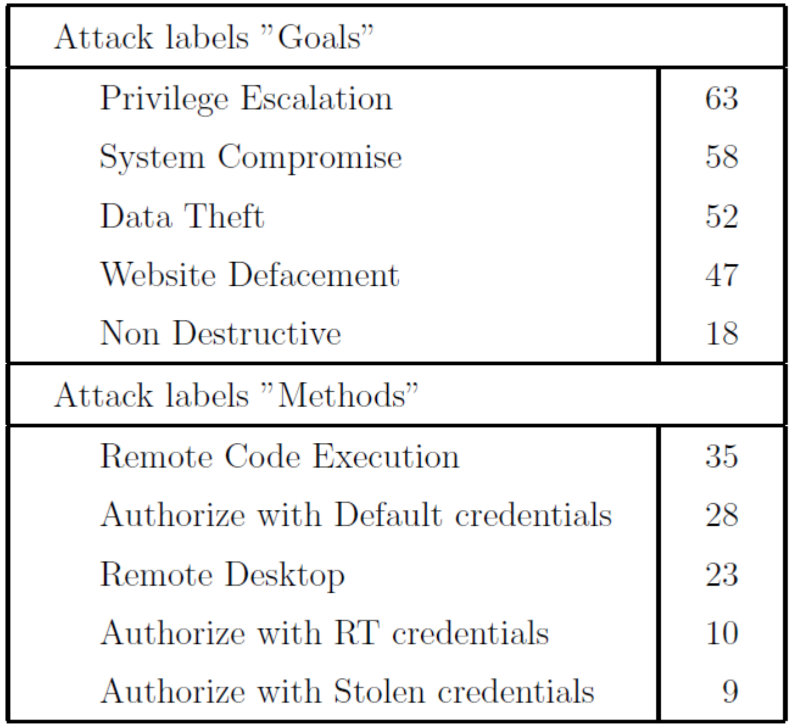
LSPR 23

LSPR 24

The LSPR23 dataset

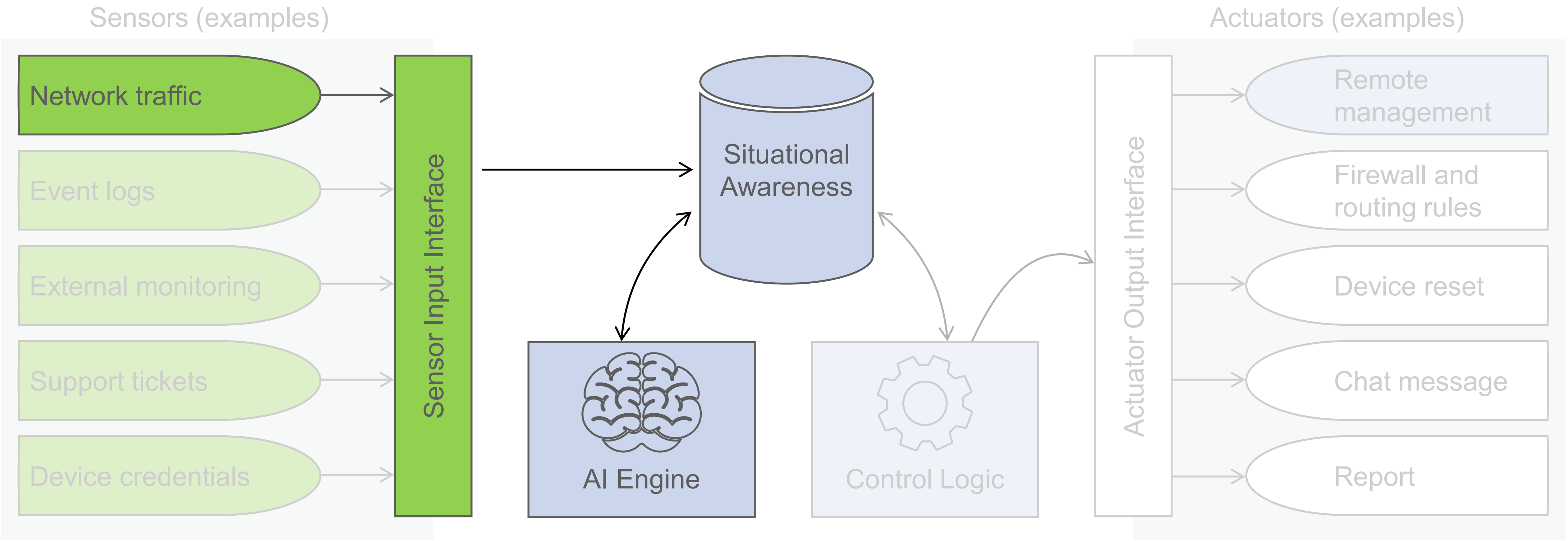


Total Flows	16,353,511	
Labels:	Source	Destination
Blue Flows	12,493,826	10,821,533
Red Flows	13,880	1,630,719
Green Flows (not scoring)	17,256	263,318
Green Flows (scoring bots)	3,415,280	136,321
External Flows	413,269	3,501,620
Benign Flows*	14,708,912	
Malicious Flows*	1,644,599	
Network segments:		
Berilia Energy Group	7,566,863	6,755,988
Berilia Airforce	4,573,708	3,849,478
Bank Of Berilia	277,705	188,984
Berilia Airforce 5G	68,369	27,649
Other	3,866,866	5,531,412



Attack labels ”Goals”	
Privilege Escalation	63
System Compromise	58
Data Theft	52
Website Defacement	47
Non Destructive	18
Attack labels ”Methods”	
Remote Code Execution	35
Authorize with Default credentials	28
Remote Desktop	23
Authorize with RT credentials	10
Authorize with Stolen credentials	9

Current status of the project



We used supervised learning to detect “Command and Control” traffic

- Initial results showed that the approach works well in some cases (when training and testing was in similar settings), but it does not generalize well

Test data

Training data		LS17 A	LS18 A
	LS17 A	0.993	0.966
	LS18 A	0.945	0.993
	LS19 A	0.743	0.928
	LS21 A	0.952	0.918

F1 scores

We used supervised learning to detect “Command and Control” traffic

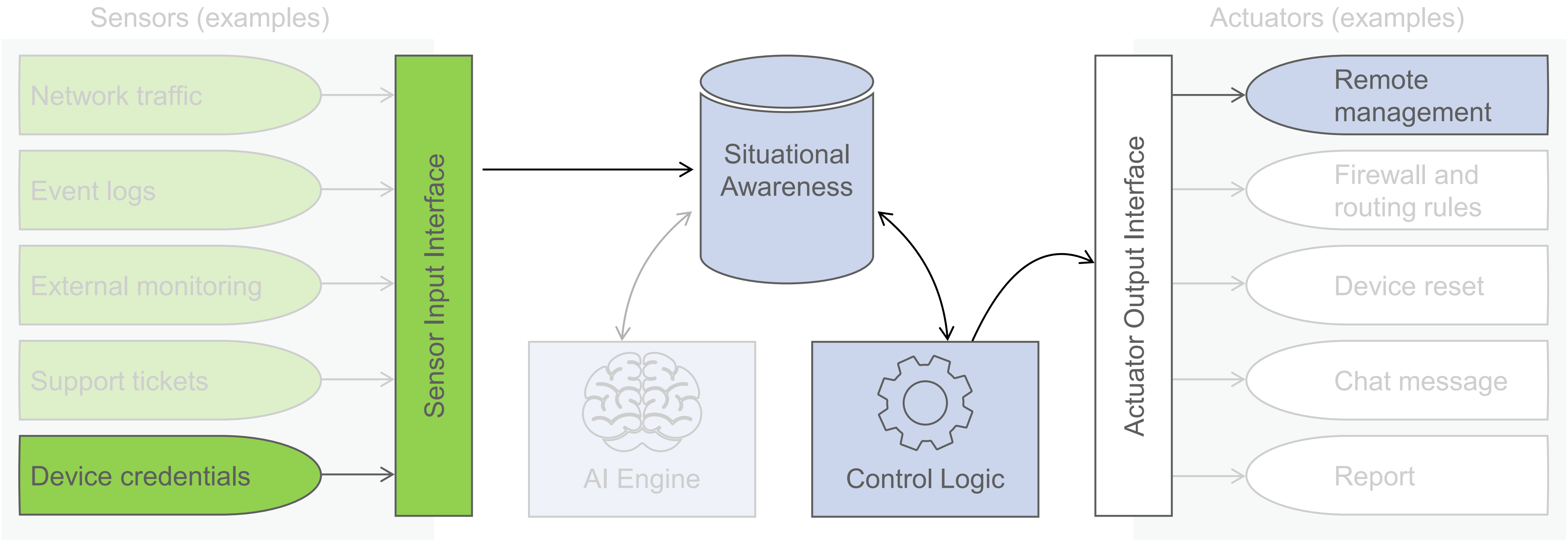
- Initial results showed that the approach works well in some cases (when training and testing was in similar settings), but it does not generalize well

Test data

Training data		LS17 A	LS18 A	LS19 A	LS21 A	LS21 B
Training data	LS17 A	0.993	0.966	0.007	0.856	0.215
	LS18 A	0.945	0.993	0.060	0.806	0.167
	LS19 A	0.743	0.928	0.791	0.351	0.000
	LS21 A	0.952	0.918	0.038	0.986	0.158

F1 scores

Current status of the project



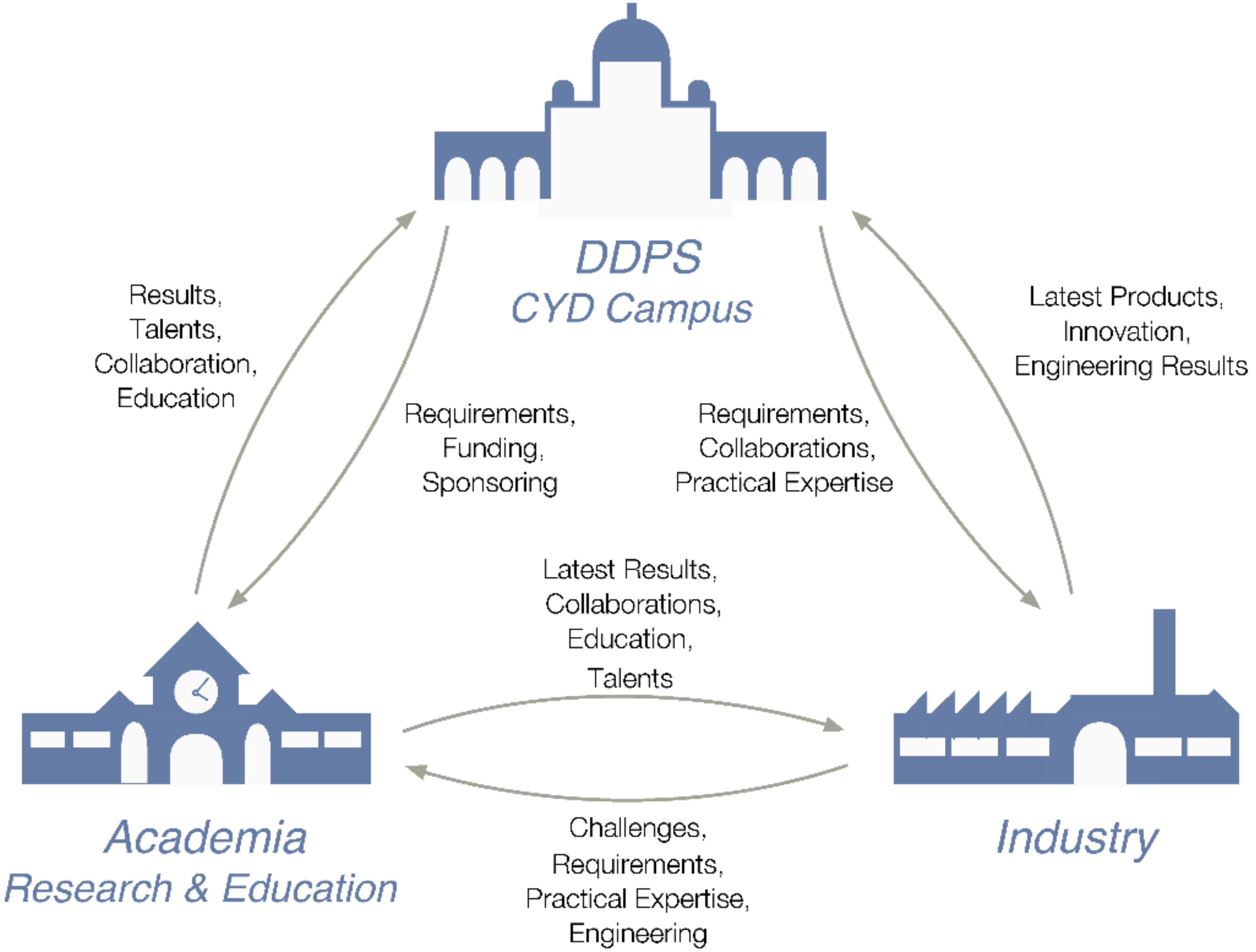
We detect and fix common misconfigurations automatically

- Currently two tools:
  - Automatically change all login credentials
  - Scan the network for nginx web servers, analyze their configuration and fix some misconfigurations automatically

Other topics we are currently (planning to) investigate

- Using Generative AI / LLMs
  - Parse support tickets
  - Generate reports
  - Generate code or configuration
- Improve existing models such that they generalize better
- Building an additional training and testing environment

The Cyber-Defence Campus connects government, academia and industry



DDPS

CYD Campus

Results,
Talents,
Collaboration,
Education

Requirements,
Funding,
Sponsoring

Requirements,
Collaborations,
Practical Expertise

Latest Products,
Innovation,
Engineering Results

Latest Results,
Collaborations,
Education,
Talents

Academia

Research & Education

Challenges,
Requirements,
Practical Expertise,
Engineering

Industry

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

CYD CYBER DEFENCE CAMPUS

CYBER DEFENCE CAMPUS ZÜRICH
CYBER DEFENCE CAMPUS THUN
CYBER DEFENCE CAMPUS LAUSANNE

Thank you for your attention!

Dr. Roland Meier

roland.meier@ar.admin.ch

cydcampus.admin.ch

cyber-defence-campus