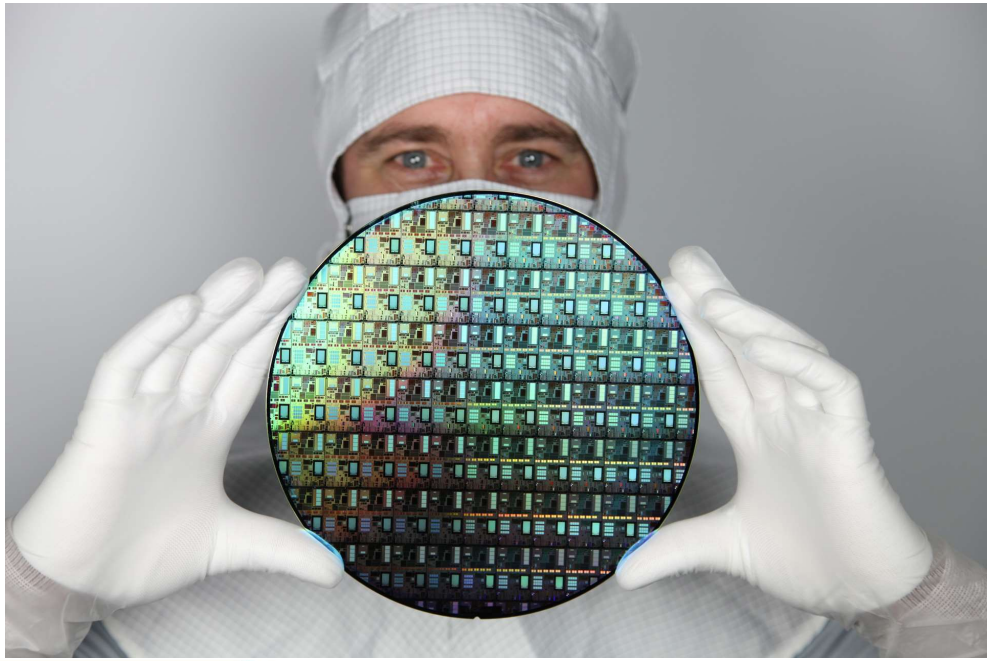
Künstliche Intelligenz zum Schutz kritischer Infrastruktur

Prof. Dr. Peter Langendörfer

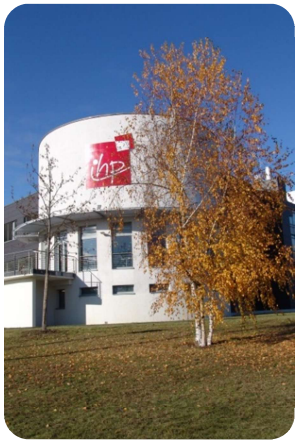
M. Aftowicz, R. Chitauro, Dr. M. Brzozowski, K. Lehniger. O. Yener, Dr. W. Alsabbagh

04.07.2024 Universität Heidelberg

IHP



IHP Impressions



Software Destroying Hardware

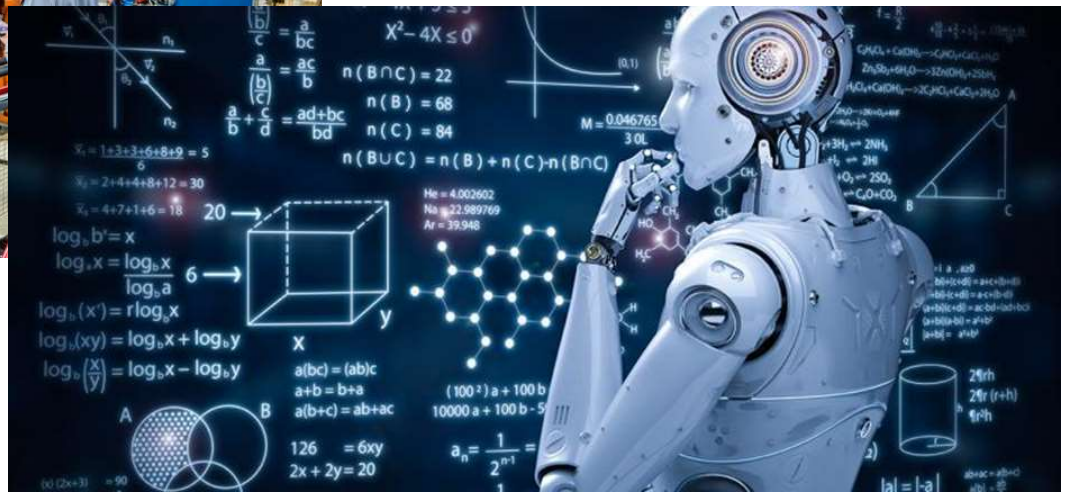
Agenda

- Einführung
  - Beispiele
  - Angreifer Klassen
  - Sicherheitsfragen
- Potentielle Opfer / Shodan
- Erkennung von Angriffen mit und ohne KI
- Erkennung von Angriffen über Anomalien
  - Wasserwerke
  - Eingebettete Systeme
- Erkennung von Jamming Angriffen
- Ein realer Fall
- Fazit

Agenda

Einführung

Einführung


$X^2 - 4X \le 0$
$n(B \cap C) = 22$
$n(B) = 68$
$n(C) = 84$
$n(B \cup C) = n(B) + n(C) - n(B \cap C)$
$\log_b b^x = x$
$\log_a x = \frac{\log_b x}{\log_b a}$
$\log_b(x^r) = r \log_b x$
$\log_b(xy) = \log_b x + \log_b y$
$\log_b\left(\frac{x}{y}\right) = \log_b x - \log_b y$
a(bc) = (ab)c
a+b = b+a
a(b+c) = ab+ac
126 = 6xy
2x + 2y = 20

GAU durch Zufall

- Slammer hat sich zunächst in den Computer eines Auftragnehmers von David-Besse eingeschleust
- So verschaffte er sich Zugang zum Unternehmensnetzwerk von David-Besse
- Er gelangte in die Prozesssteuerungssysteme des Reaktors, da das Prozesssteuerungssystem mit dem öffentlichen Unternehmensnetzwerk verbunden war
- Der Wurm sperrte die Mitarbeiter der Reaktoranlage vom Sicherheitsparameter-Anzeigesystem aus
- So konnten die Mitarbeiter beispielsweise die Kerntemperatursensoren der Anlage nicht mehr überwachen, was in einem Kernkraftwerk ein entscheidendes Sicherheitsrisiko darstellt

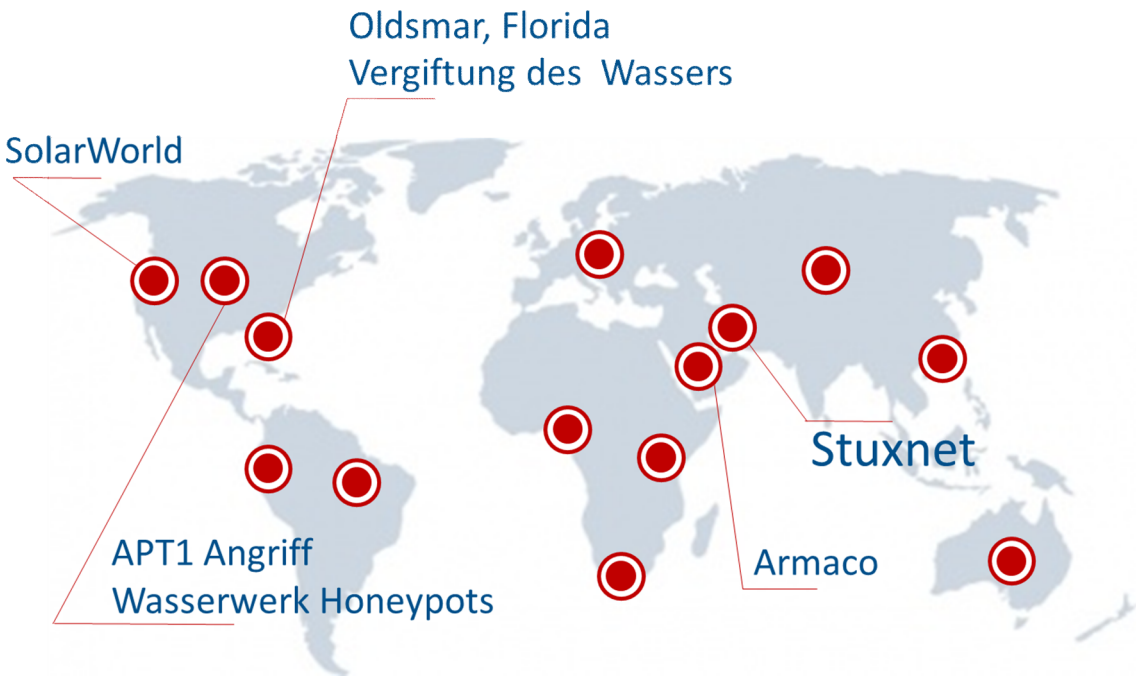
Immer grüne Welle

- Verwendung von unverschlüsselter drahtloser Kommunikation
- Standard-Passwörter und Debug-Ports
- In einigen Fällen unauthentifiziertes Schalten für Notdienste aktiviert

Hacking automation control systems for dummies

- Zahl der ungeschützten PLCs stieg von 2013 bis 2015 um 300 Prozent 28000 im Jahr 2015 gefunden
- Angriff: Code runterladen, verändern und wieder hochladen
- Holen Sie sich Ihr Toolkit unter: https://github.com/SCADACS/PLCinject

Bekannte Industrielle Cyber-Angriffe gegen Kritis



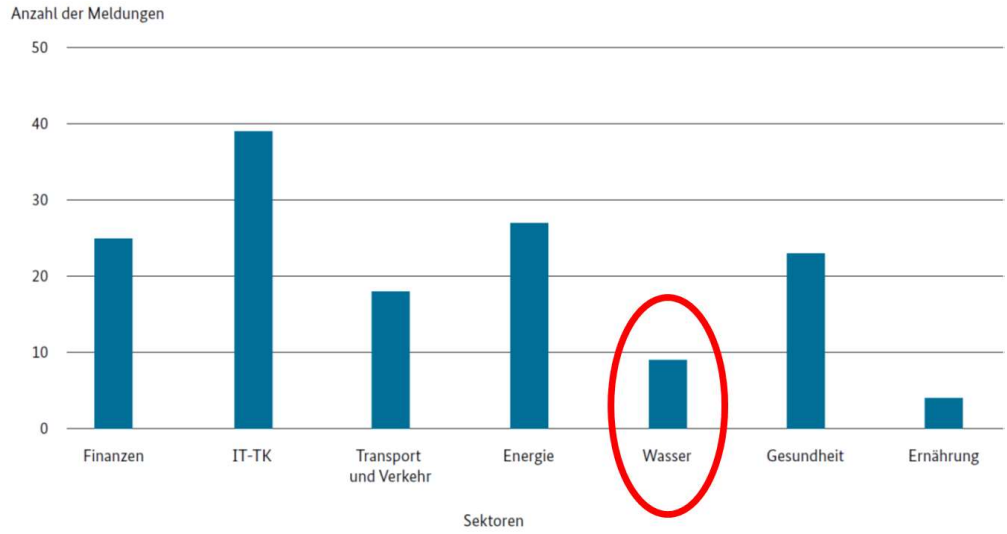
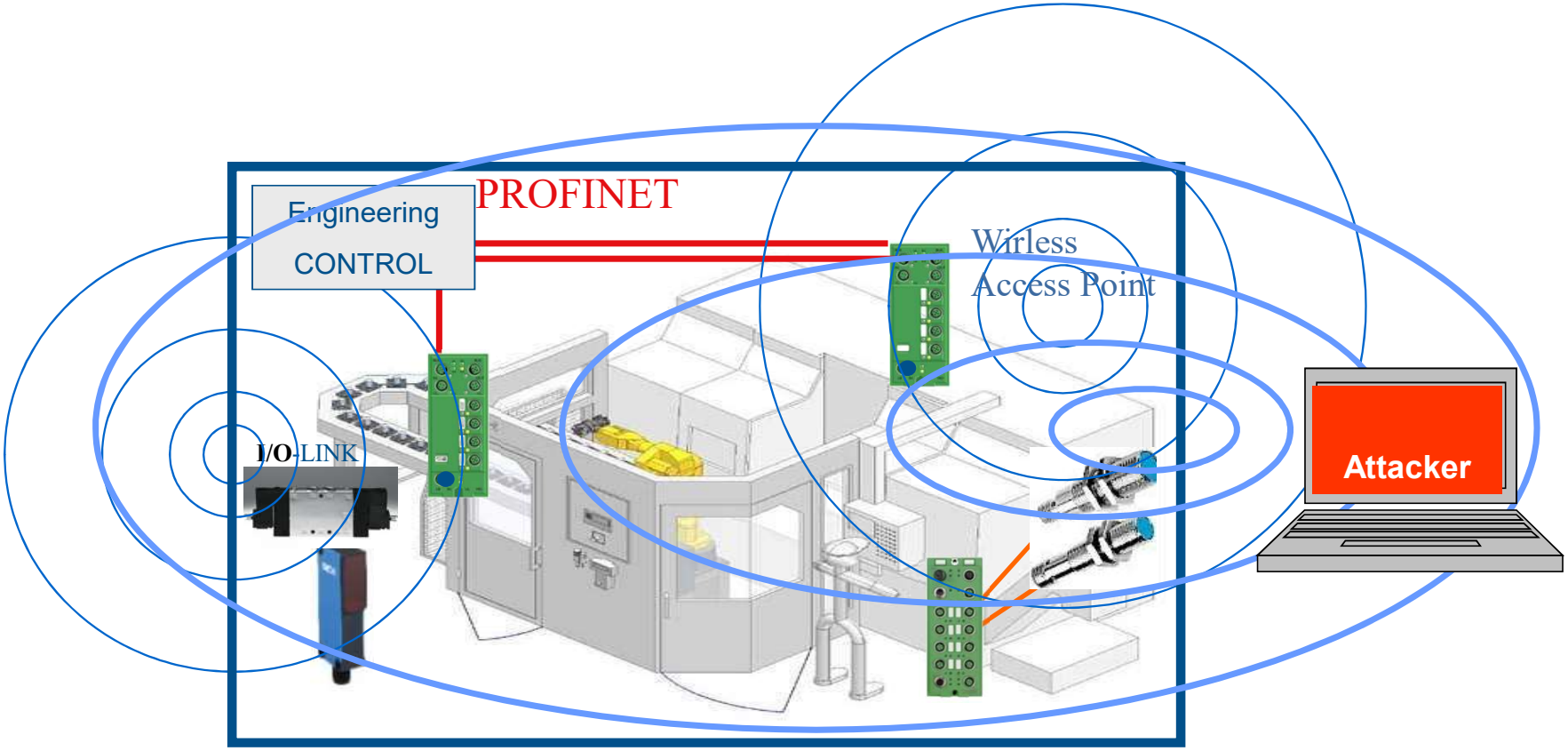


Abbildung 02 Meldeaufkommen von KRITIS-Betreibern (freiwillige und verpflichtende Meldungen nach § 8b BSIG) im Berichtszeitraum vom 01.06.2017 bis 31.05.2018

Industrie Automatisierung: Mögliche Angriffe



Angreifer Klassen

- Einzelpersonen, z. B. Teenager/Studenten; Ziel: Spaß
- Kriminelle Identität unbekannt; Ziel: monetäre Vorteile z. B. durch Erpressung
- Nationen; Ziele: Macht, Spionage, Rache, Cyber-Kriegsführung .....
  - Stuxnet, teilweise vom Iran nachgebaut, wurde zur Zerstörung der Server der nationalen Öl- und Erdgasgesellschaft Saudi-Arabiens eingesetzt (laut Ian Bremmer)Das
  - Angriff durch russische Hacker auf SCADA Systeme mit der Malware BlackEnergy (laut US-Heimatschutzministerium)
  - Das SCADA-System von Schneider Electric wurde Berichten zufolge von Chinesen gehackt.

IT-Sicherheitsfragen

Hauptproblem: für den Inselbetrieb konzipierten Systeme sind nun miteinander verbunden:

- Inflight Entertainment
- On-Board-Einheiten (Fahrzeugunterhaltungssysteme)
- Mit dem Internet verbundene Automatisierungssteuerungssysteme
- Gast-Login

Angemessene Sicherheitsvorkehrungen fehlen:

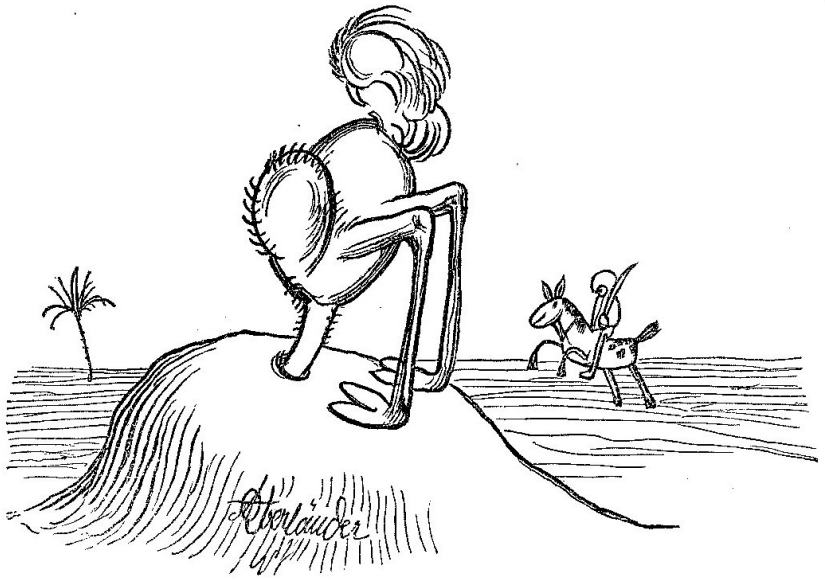
- Keine Firewall,
- keine IDS-Unterstützung
- keine Datenintegritätsprüfungen,
- keine Authentifizierung z.B. CAN-Bus bietet nicht einmal Platz in seinen Paketen, um solche Funktionen nachzurüsten

Agenda

Potentielle Opfer / Shodan

Was geht mich das an?

Zu klein um angegriffen zu werden

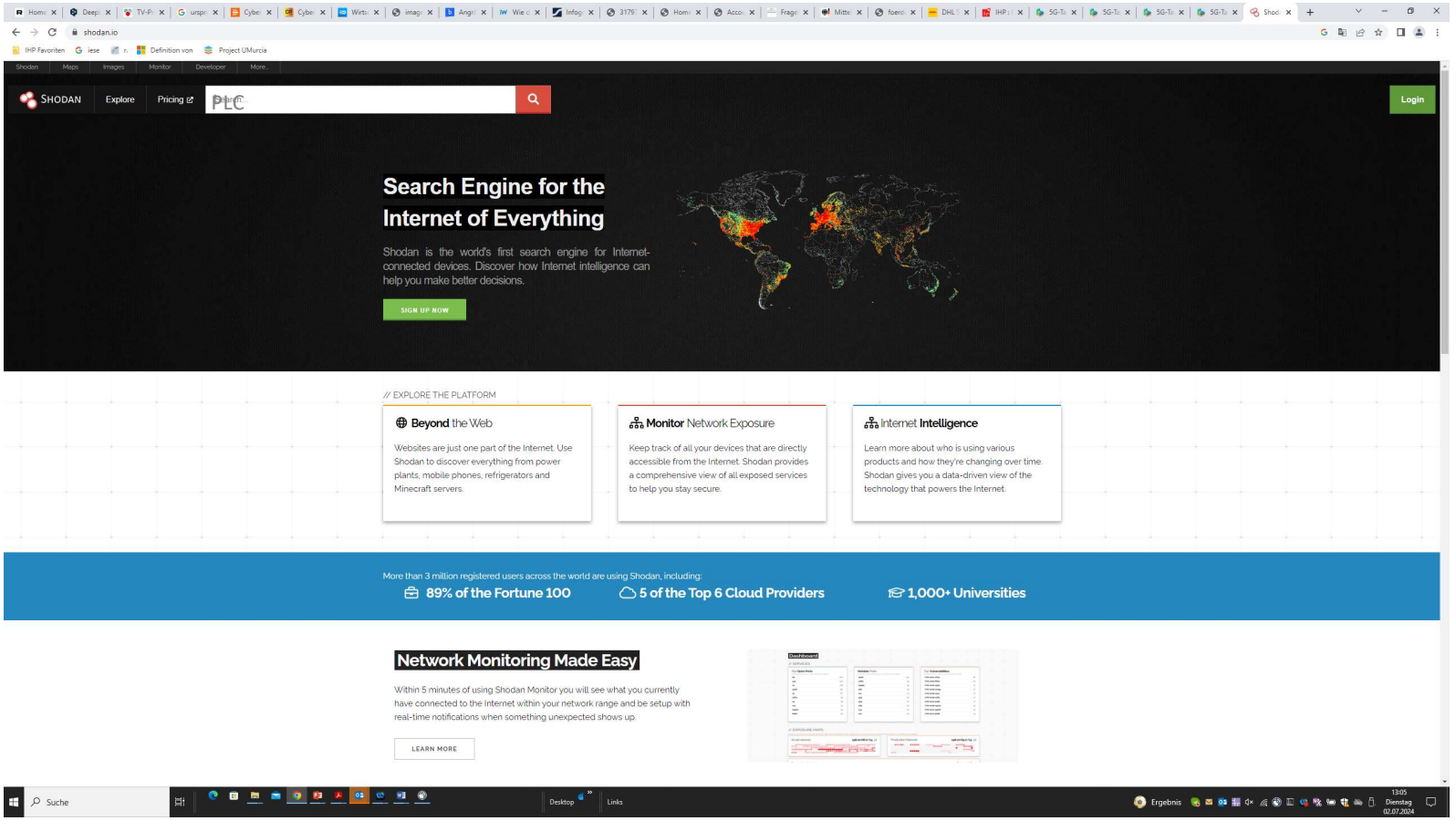


Shodan: Google for the Internet of Things

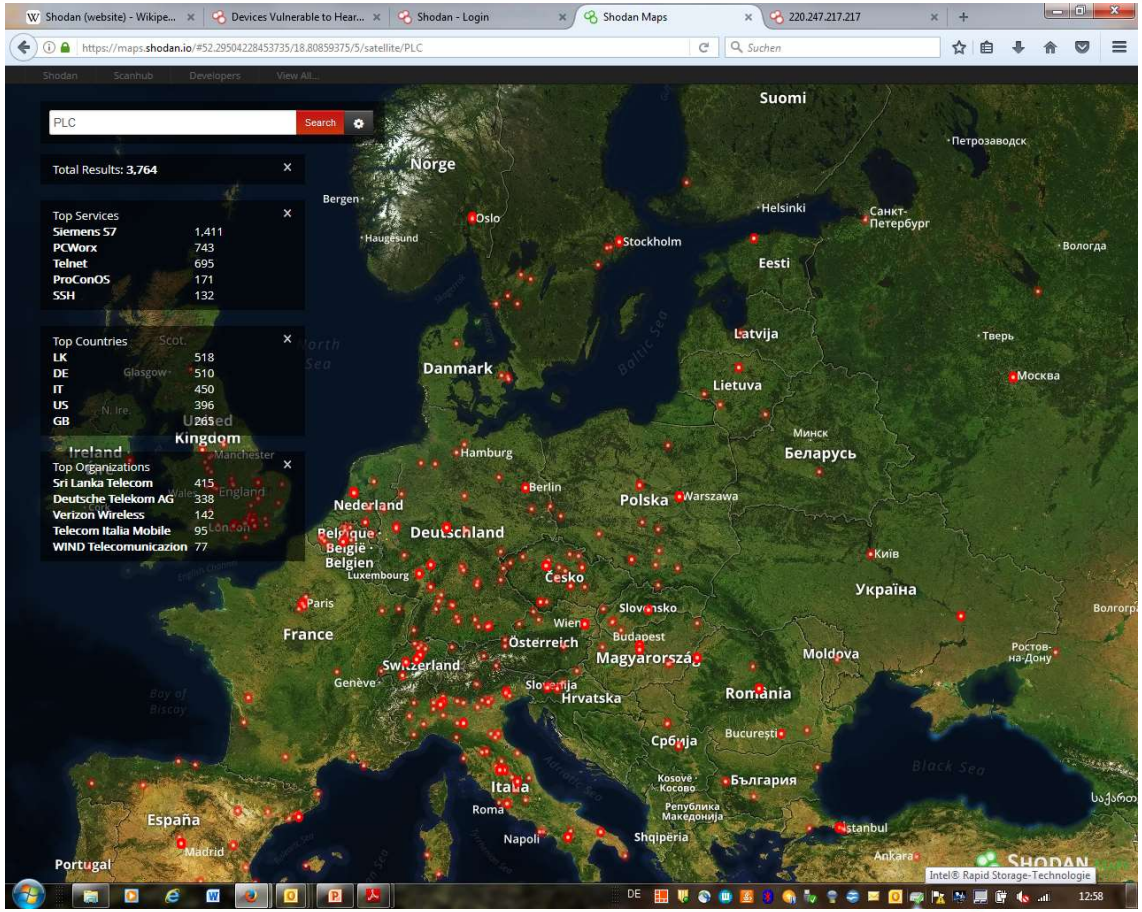
Shodan: Google für das Internet der Dinge

- Suchmaschine kann verwendet werden, um folgendes zu finden:
  - Router mit offenen Hintertüren,
  - ungesicherte Webcams und
  - industrielle Kontrollsysteme, die noch Standardpasswörter verwenden
  - Server mit fehlerhaften Implementierungen, z. B. Heartbleet
- Die öffentliche Website ist nur ein kleiner Teil dessen, was Shodan bietet
- Spielwiese für Hacker und Terroristen ODER nützliches Werkzeug für Unternehmen???
- Unternehmenskunden können Echtzeit-Zugang zu allen gesammelten Daten erwerben.
- Ein Unternehmen kann Shodan zum Beispiel nutzen, um seine eigenen Netzwerke zu durchsuchen.

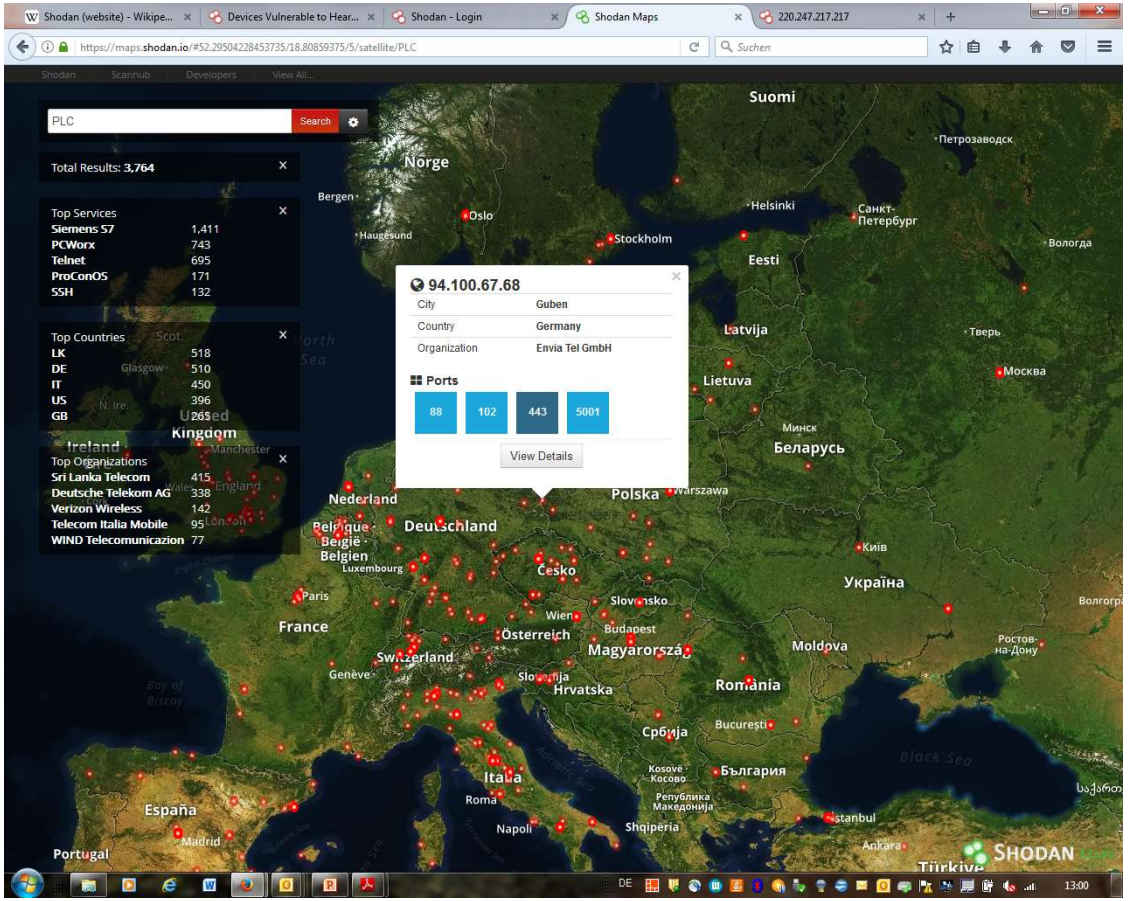
Shodan suche nach SPS/PLC



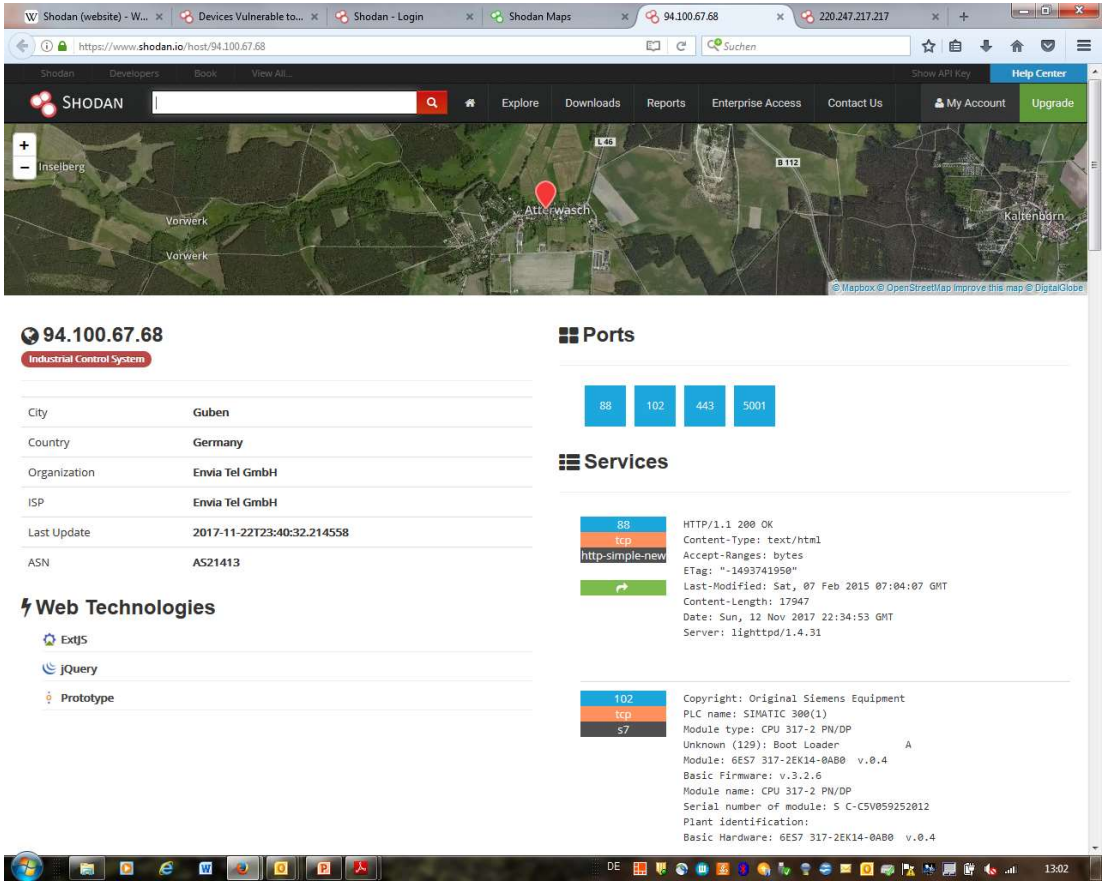
Shodan Suchergebnisse für SPS zoom in



Shodan Suchergebnisse für SPS zoom in



Shodan Suchergebnisse für SPS zoom in



Suchergebnisse für Shodan etc.

Table 1.1: The number of Internet-accessible PLCs on March 22, 2023

Protocol	Port	*Censys*	*Shodan*	*ZoomEye*	*Ditecting*
Modbus	502/TCP	35,018	61,902	31,706	14,173
Siemens S7	102/TCP	6,804	56,010	40,099	2464
DNP3	20000/TCP	597	962,297	14,750,352	367
BACnet	47808/TCP	15,514	31,157	56,767	11,750
Niagara Fox	1911/TCP	25,398	80,902	262,111	26,634
Ethernet/IP	44818/TCP	Not Available	65,402	27,637	1971
Phonix/PCWorx	1962/TCP	Not Available	48,924	316,365	168
Codesys	2455/TCP	Not Available	37,099	241,809	1298

Agenda

Erkennung von Angriffen mit und ohne KI

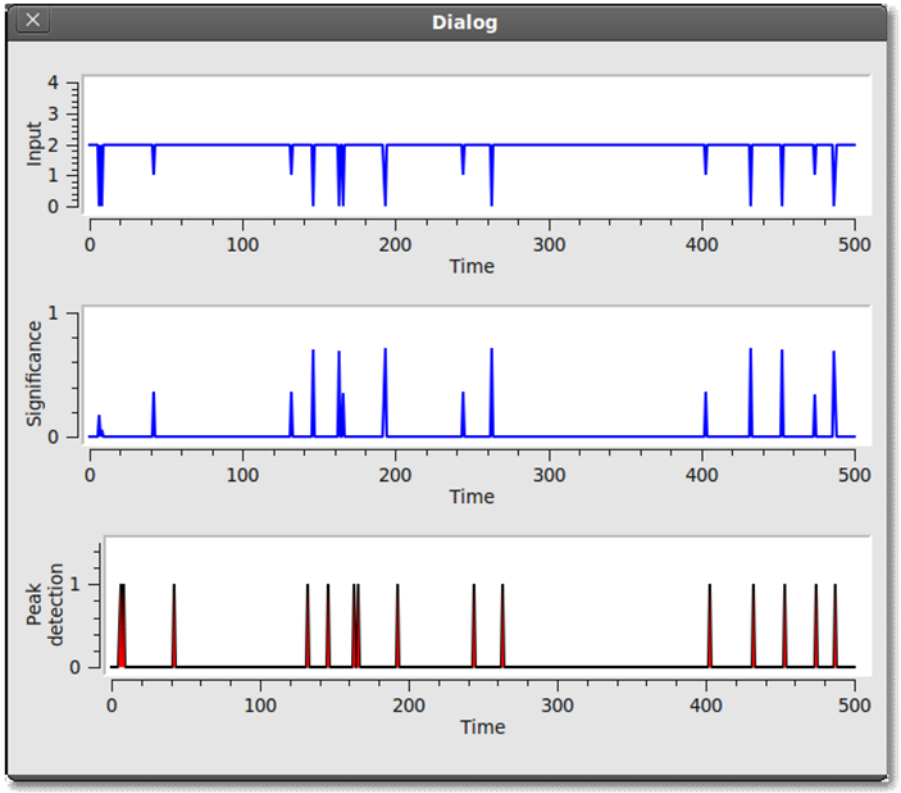
Hintergrund: Trinkwasser Pipeline nach Frankfurt

Hintergrund: Trinkwasser Pipeline nach Frankfurt

Messungen erfolgen alle 30 Sekunden und werden dann übertragen

Location	Distance to next substation	Parameter	Unit
Waterworks Briesen	~1800 m	flow rate	(m^3/h)
		pressure, outlet	(bar)
		position of butterfly valve	indication
Briesen (protection on pipe bursts)	<~5800m	pressure pipe A(1)	(bar)
		pressure pipe N(2)	(bar)
		position of butterfly valve	indication
Jacobsdorf (protection on pipe bursts)	~4800m	flow rate	(m^3/h)
		quantity	(m^3)
		pressure pipe A(1)	(bar)
		pressure pipe N(2)	(bar)
		position of butterfly valve	indication
Pilgram/Pagram (protection on pipe bursts)	~2500m	pressure pipe A(1)	(bar)
		pressure pipe N(2)	(bar)
		position of butterfly valve	indication
Reservoir	0m	flow rate	(m^3/h)
		quantity intake from waterworks with negativ back flow	(m^3)
		quantity intake from waterworks	(m^3)

Anomaly Erkennung durch Signifikanz Analyse

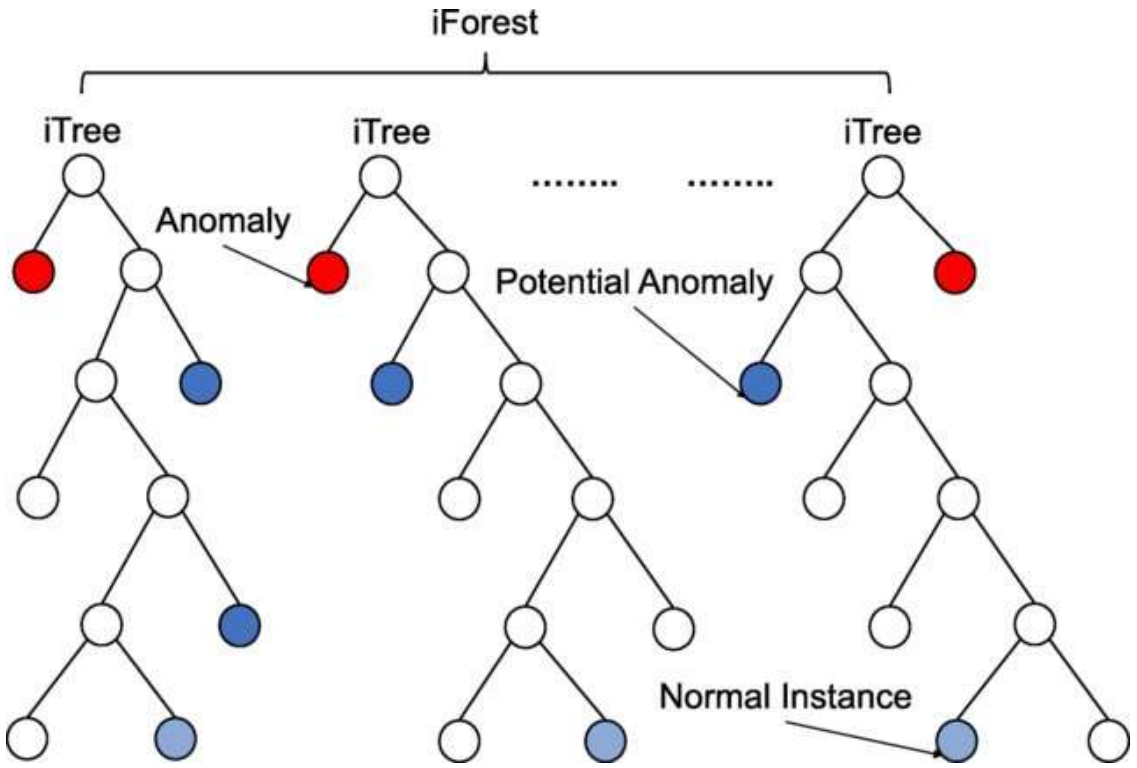


Ausbleibende Nachrichten lösen einen Alam aus

Signifikanzanalyse:

$$\sigma_m = \sqrt{(\frac{1}{n}\sum_{i=1}^{n} x_i^2) - \bar{x}^2} = \sqrt{(\frac{1}{n}\sum_{i=1}^{n} x_i^2) - (\frac{1}{n}\sum_{i=1}^{n} x_i)^2}$$

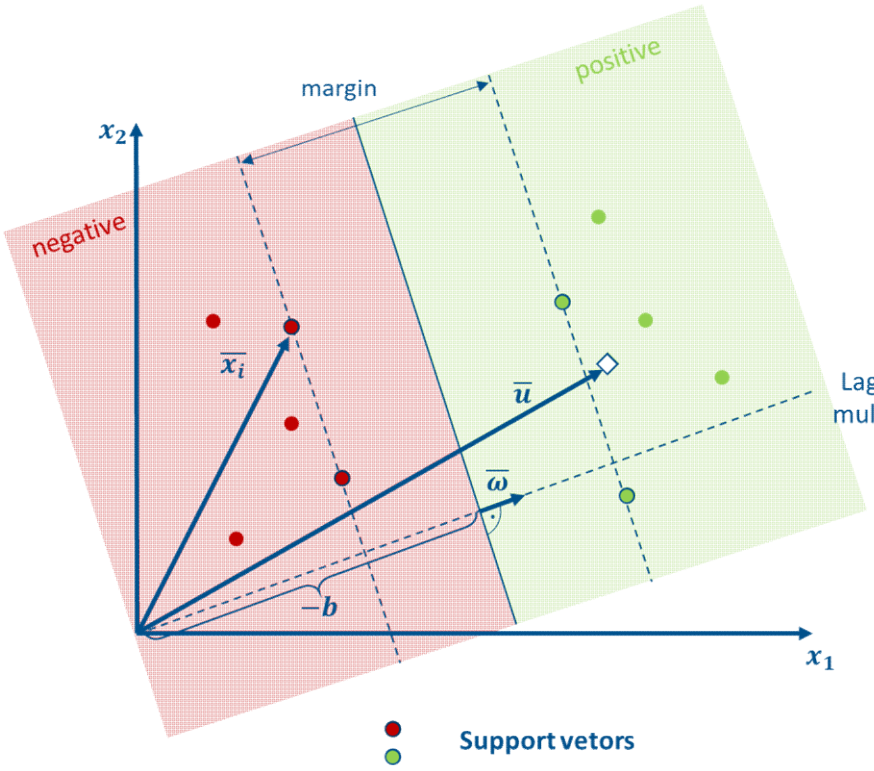
Isolation Forrest



- Länge des Pfades durch den Baum kennzeichnet normale bzw. und anormale Daten
- Geringe Speicheranforderungen
- Geringe Anforderungen an die Rechenleistung
- Training ohne Anomalien möglich
- Gut geeignet für "billige" Geräte

Regaya, Y., Fadli, F. & Amira, A. Point-Denoise: Unsupervised outlier detection for 3D point clouds enhancement. Multimed Tools Appl 80, 28161–28177 (2021).

Support Vector Machines



$$\sum_i \bar{\omega} \cdot \bar{u} + b \geq 0$$

$$\sum_i \overbrace{\alpha_i y_i \bar{x}_i}^{\bar{\omega}} \cdot \bar{u} + b \geq 0$$

Lagrange multipliers

Support vetors

labels

$$\sum_i \alpha_i y_i K(\bar{x}_i, \bar{u}) + b \geq 0$$

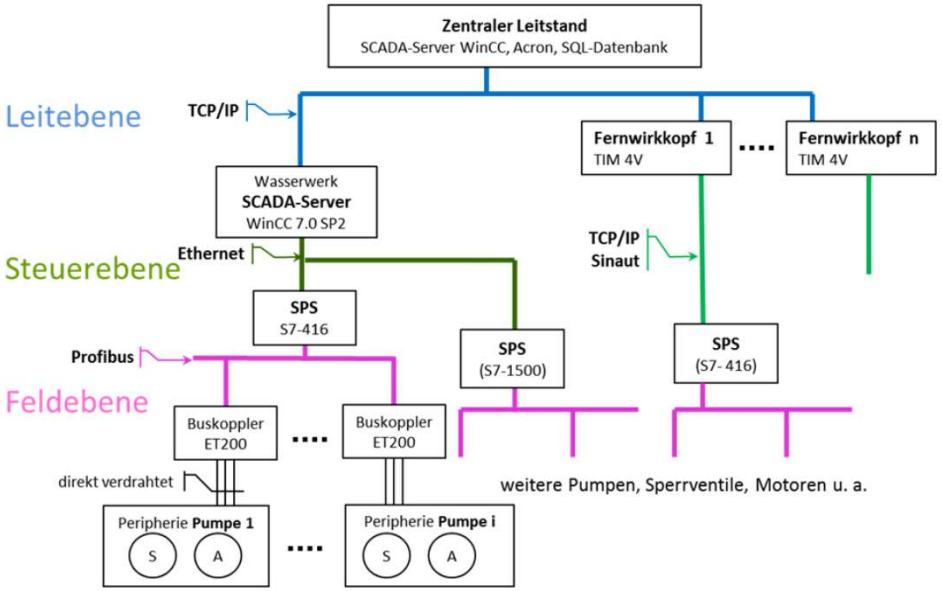
- Erzeugt eine Hyperebene, die normale und anormale Daten voneinander trennt
- Geringe Speicheranforderungen
- Geringe Anforderungen an die Rechenleistung
- Training ohne Anomalien möglich
- Gut geeignet für "billige" Geräte

Agenda

Erkennung von Angriffen über Anomalien

KI-basierte Anomalie-Erkennung in Wasserwerken

- IT und OT wurden früher getrennt betrieben
- Pumpen/Sensoren sind IoT Geräte
- Das Angriffsrisiko ist aktuell größer als jemals zuvor
- Typische IDS analysieren IT Netzwerke
- Wichtig:
  - Analyse von Sensoren, Geräten aus dem OT Bereich
  - Kombination mit “Standard” IDS



KI-basierte Anomalie-Erkennung in Wasserwerken

- SWAT Data set
  - 11 Tage Datenerfassung
  - Testbed Betrieb 24/7
  - ersten sieben Tagen lief das Testbed unter normalen Bedingungen
  - vier Tagen der Datenerfassung wurden Angriffe gestartet.
- Systematische Erstellung der Angriffe anhand eines Angriffsmodells, das die Absichten der Angreifer berücksichtigt.
- 36 verschiedene Angriffe auf das SWaT-Testbed gestartet, es wurden hauptsächlich Sensoren angegriffen, indem falsche Daten an die SPSen gesendet wurden (z.B. raw water tank level, pH Analyzer, Flow rate sensor):
- Der Netzverkehr wurde mit handelsüblicher Ausrüstung von Check Point R Software Technologies Ltd2 erfasst.
- Kommunikation zwischen dem SCADA-System und den PLCs.
- Angriffe:
  - Abfangen der Pakete während der Kommunikation zwischen dem SCADA-System und den SPSen
  - Veränderung der Netzwerkpakete so, dass sie die gefälschten Werte der Sensoren widerspiegeln

KI-basierte Anomalie-Erkennung in Wasserwerken

- Labelling der physikalischen Eigenschaften
  - Jedes Datenelement, das einem Sensor oder einem Aktor entspricht, wurde einzeln in einer CSV-Datei erfasst.
  - Jede CSV-Datei enthält den Servernamen, den Sensornamen, den Wert zu diesem Zeitpunkt, den Zeitstempel, die
  - Alle Daten wurden dann in einer einzigen CSV-Datei zusammengefasst.
  - Anhand der Angriffsprotokolle wurden die Daten auf der Grundlage der Start- und Endzeiten der Angriffe manuell gelabelled
- Labelling des Netzwerkverkehrs
  - Aufteilung der Netzwerkdaten in mehrere CSV-Dateien mit einer Zeilenbegrenzung von 500.000
  - Erfassung im Sekundentakt führt zu Überschneidungen, wenn mehrere Zeilen eine andere Aktivität widerspiegeln, aber denselben Zeitstempel tragen.
  - Labelling mit Hilfe der Angriffsprotokolle anhand analog zu den physikalischen Eigenschaften
- In KISS_KI ging es hauptsächlich um OT Verfälschungen.
- Die Idee war IT Angriffe mit aktuellen IT Systemen zu finden (z.B. IDS) und nach einem erkannten IT Angriff wurden KI für OT Daten als sensibler konfiguriert

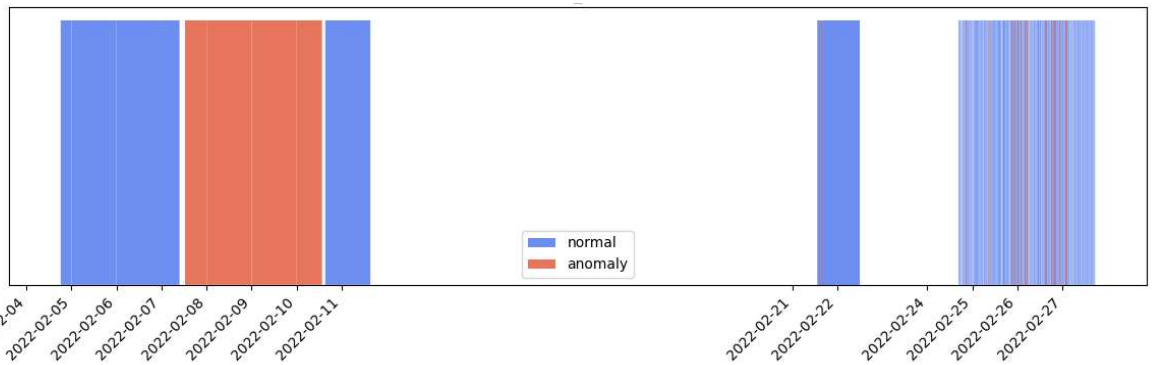
Anomalie Erkennung in industriellen Kontrollsystemen

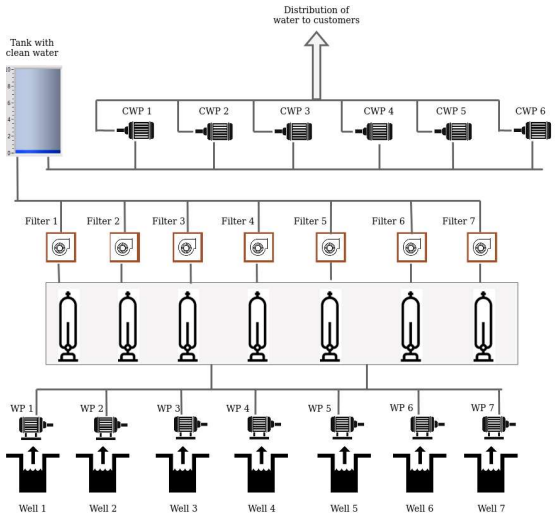
- Validierung der Trainingsdaten mit Open-Source-Datensätzen

SWaT	precision	recall	f1_score
One Class SVM	0.94	0.65	0.77
Isolation Forest	0.95	0.62	0.75

- One Class SVM Hardware ohne Leistungseinbußen implementiert

One Class SVM	ARM 667 MHz	FPGA 150 MHz
Anomaly detection	1 ms	21 μs





Industrie Kooperation:

- Wasserversorgungsgesellschaft **DNWAB**
- Modellierung der Wasseraufbereitung **Aucoteam**
- Modellierung der Wasserverteilung **Sieker**
- IT Sicherheit **@yet**

RoP Angriffe

- Pufferüberläufe sind seit Jahrzehnten ein Problem, insbesondere durch die Nutzung unsicherer Programmiersprachen wie C oder C++.
- Bei einem Pufferüberlauf werden mehr Daten in einen Puffer kopiert als dieser fassen kann.
- Dadurch können sensible Daten überschrieben werden.
- Ein beliebtes Ziel sind dabei die Rücksprungadressen von Funktionen.
- der Angreifer verändert damit den Kontrollfluss das Verhalten eines Programms beeinflussen.
- Oftmals werden dafür kleine Programmschnipsel (Gadgets) verwendet, die mit einem Rücksprung (ret Instruktion) enden.
- Ist ein angegriffenes Programm groß genug lassen sich darin genug Gadgets finden, um beliebiges Verhalten zu erzeugen.

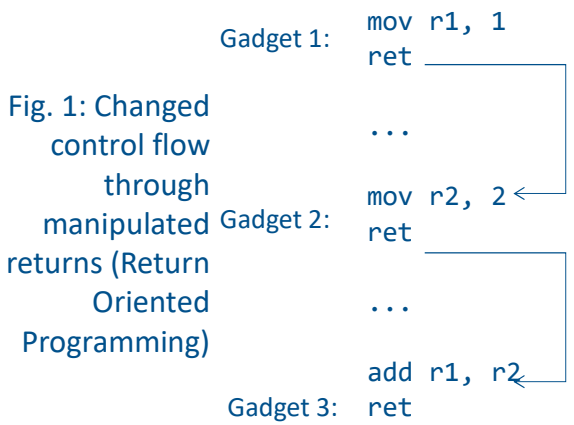
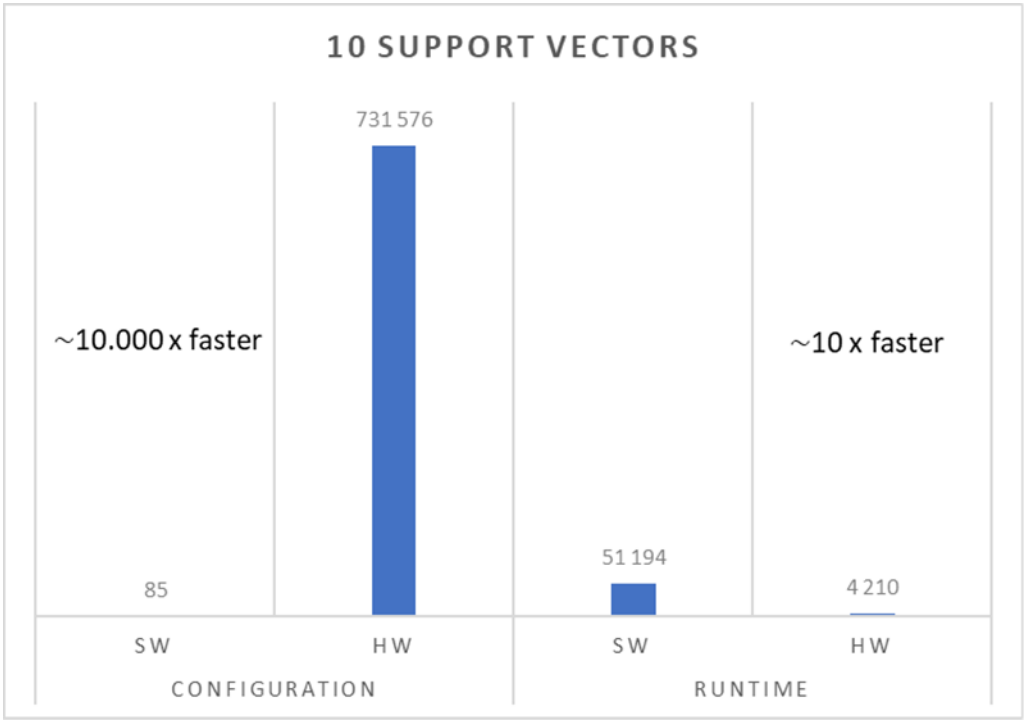


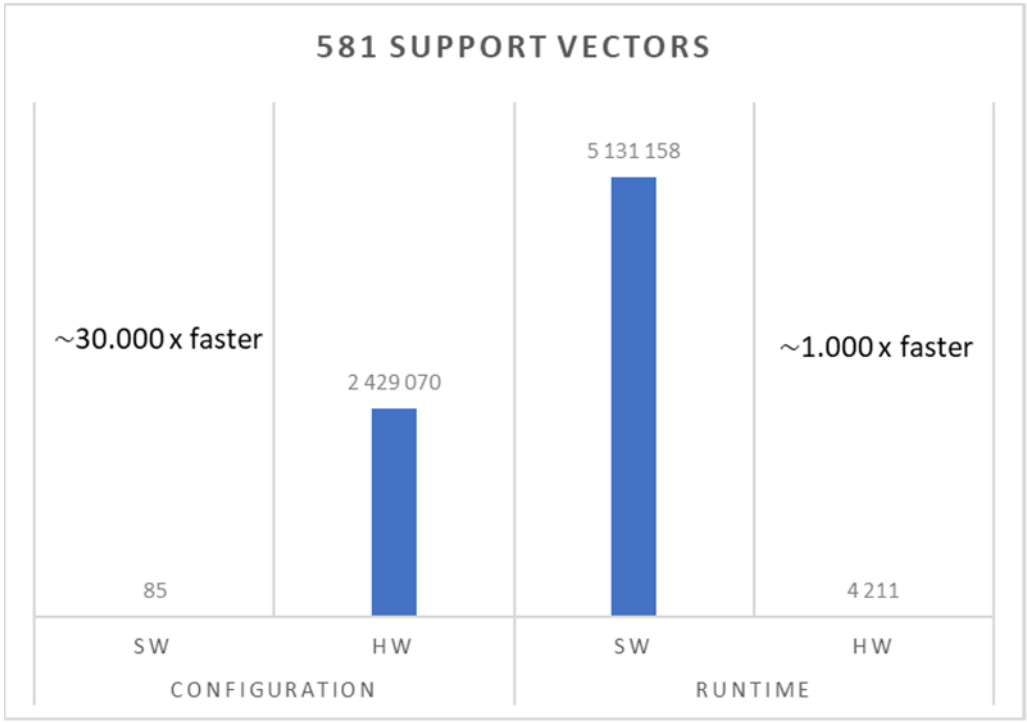
Fig. 1: Changed control flow through manipulated returns (Return Oriented Programming)

Fig 2: Attacker writes across buffer boundaries and manipulates underlying data

K. Lehniger and P. Langendörfer, "Window Canaries: Re-thinking Stack Canaries for Architectures with Register Windows," in IEEE Transactions on Dependable and Secure Computing, doi: 10.1109/TDSC.2022.3230748.

Software SVM vs. Hardware SVM



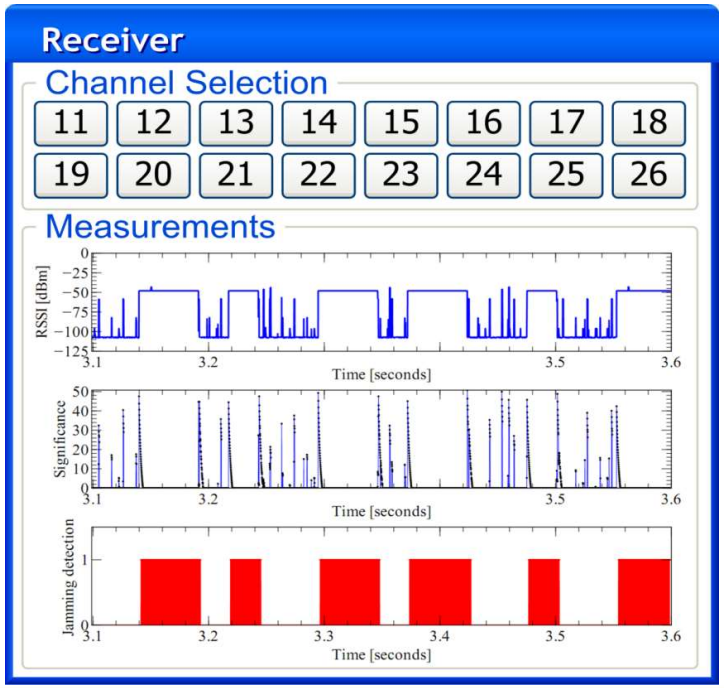


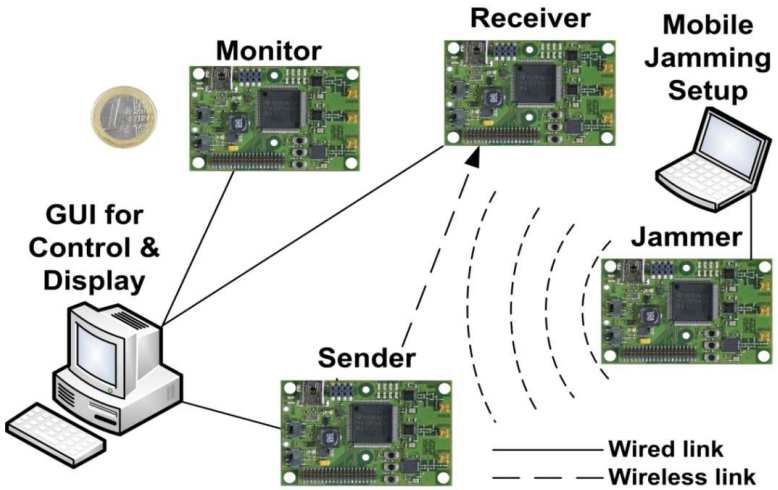
Agenda

Erkennung von Jamming Angriffen

Jamming Erkennung ohne KI

Jamming Erkennung ohne Festlegung von Schwellwerten für RSSI





MSP430 microprocessor

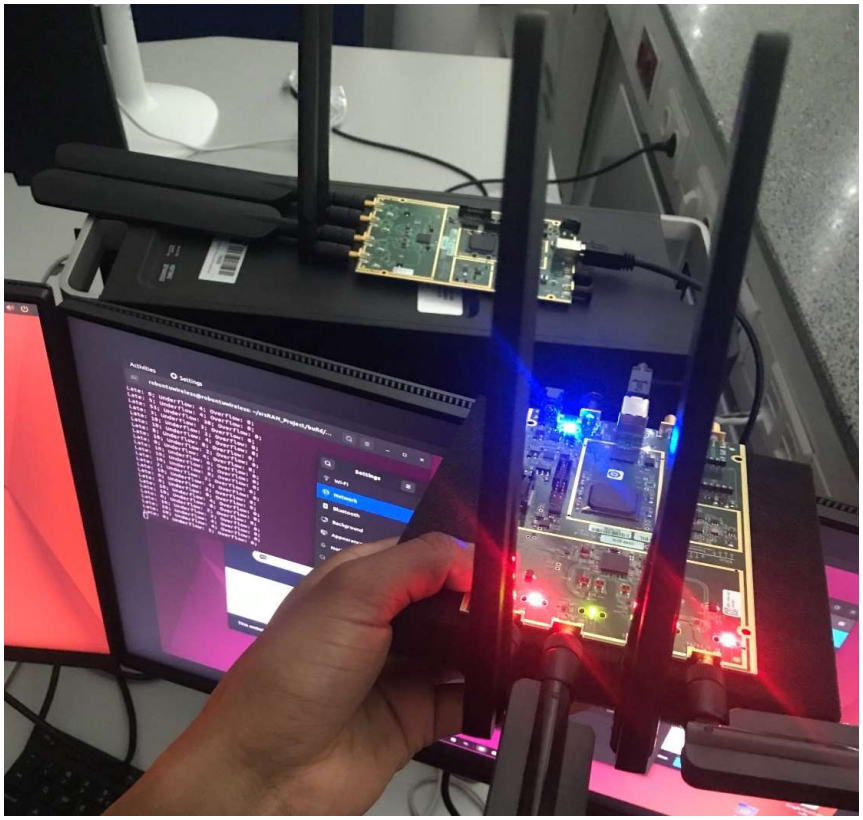
- **422 Bytes of memory für die** Signifikanz Analyse inklusive aller Parameters und des Codes
- Erkennungsrate **>95 Prozent**
- **3 Typen von J**ammern können erkannt werden: konstante, periodische, zufällige

Jamming Erkennung

- KPI Analyse:
  - SINR (Signal-to-Interference-plus-Noise Ratio),
  - BER (Bit Error Rate),
  - MCS (Modulation and Coding Scheme), und
  - CQI (Channel Quality Indicator).

- Spektrogramm Analysis:
  - Ein Machine Learning Modell lernt den Unterschied zwischen Spektrogrammbildern mit und ohne Störeinflüsse

Jamming Erkennung

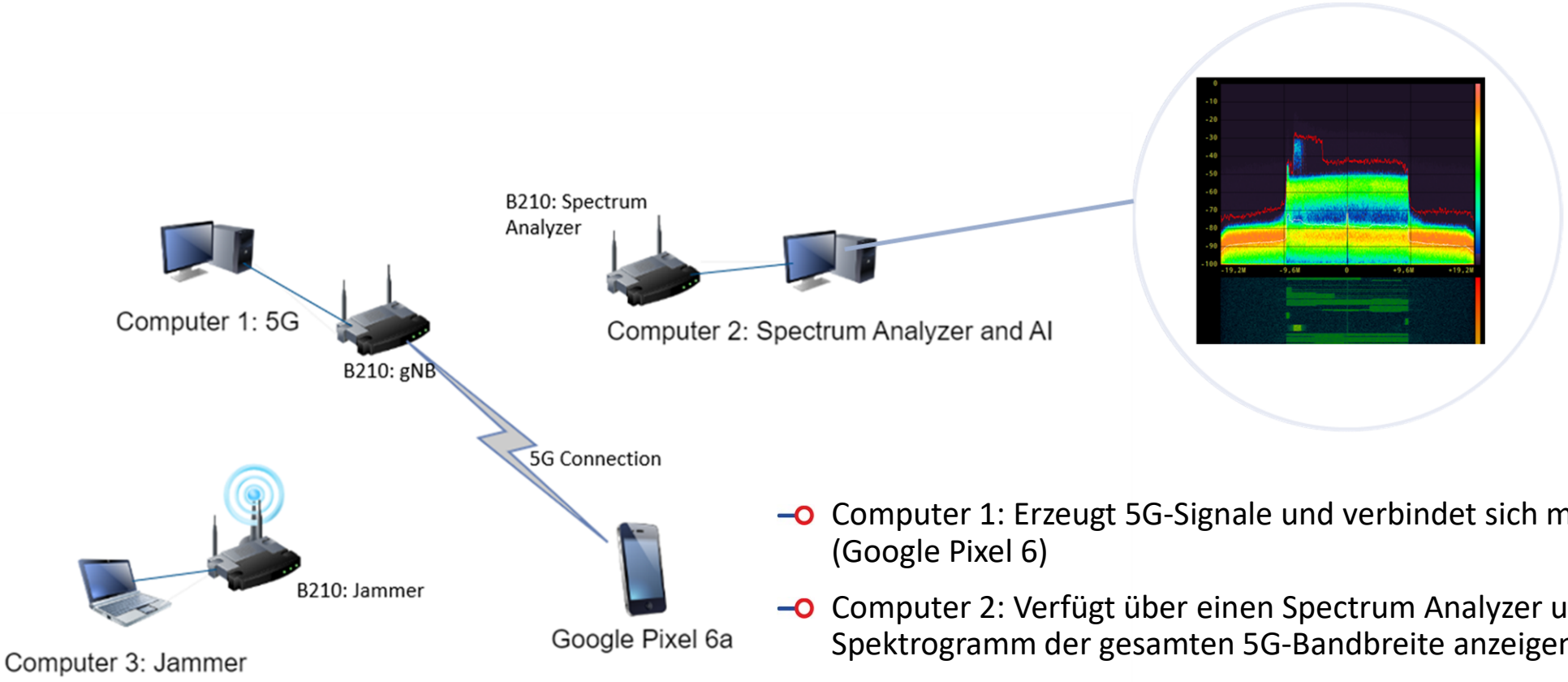


USRP B210 SDR transmitting 5G signals as a gNB.

- Daten sind für die Erstellung von Modellen für maschinelles Lernen unerlässlich
- Wir haben unsere Daten von einem End-to-End-5G-Testbed gesammelt, das Signale Over The Air (OTA) überträgt
- Wir verwendeten das srsRAN-Projekt [1] zur Bereitstellung von RAN-Funktionen und Open5gs [2] für Kernnetzfunktionen
- USRP B210 SDR wurde als Basisstation (gNB) und ein Google Pixel 6a Telefon als UE verwendet

[1] srsRAN Project Documentation — srsRAN Project documentation
[2] open5gs.org

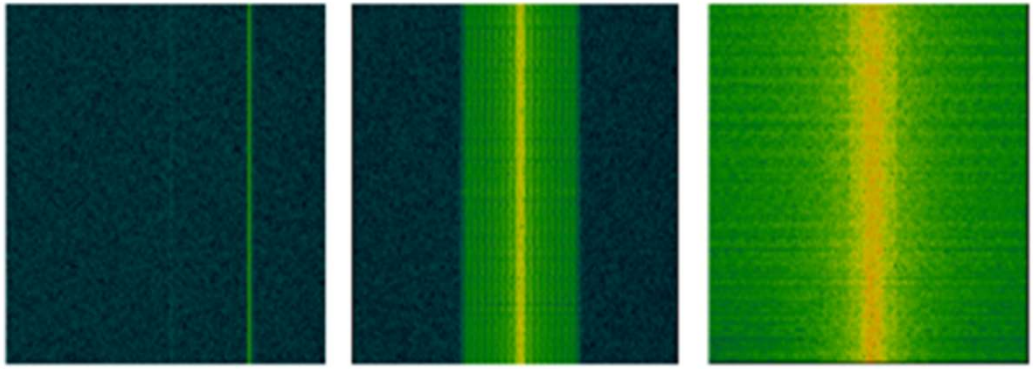
Jamming Erkennung



- Computer 1: Erzeugt 5G-Signale und verbindet sich mit dem UE (Google Pixel 6)
- Computer 2: Verfügt über einen Spectrum Analyzer und kann das Spektrogramm der gesamten 5G-Bandbreite anzeigen.
- Computer 3: Kann Jamming-Signale erzeugen.

Jamming Erkennung

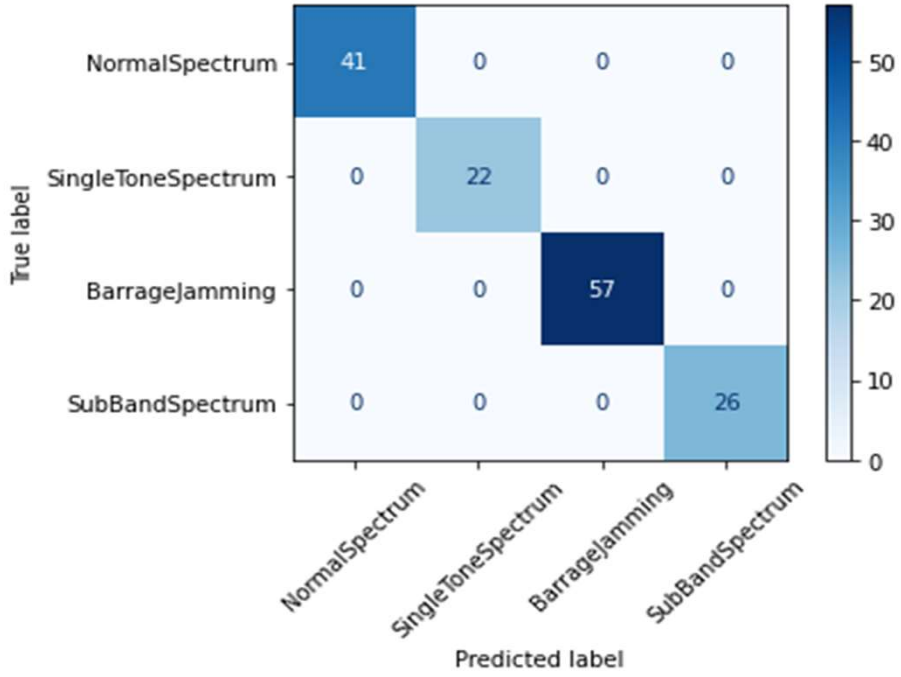
- Single-Tone Jammer: Überträgt ein Störsignal auf einer einzigen Frequenz innerhalb der 5G-Bandbreite
- Sub-Band-Jammer: Überträgt Störsignale, die einen Teil der 5G-Bandbreite abdecken
- Barrage Jammer: Überträgt Störsignale, die die gesamte 5G-Bandbreite abdecken.
- Spektrogrammbilder jedes Störungstyps werden erfasst, gespeichert und nach ihrem Typ gruppiert. Diese werden als Trainingsdaten verwendet.
- Auch die gesammelten Störungsdaten - mit Metriken wie Zeitstempel, Störleistung, Dauer und gegebenenfalls Zentralfrequenz und Bandbreite - werden aufgezeichnet.



Jamming Erkennung

- Ausnutzen des Reifegrad von Bildklassifizierungsalgorithmen
- CNN-Architektur als Vorlage, hier EfficientNetV2B0
- Anpassung der Ausgabeschicht an unser Klassifizierungsszenario
- Transfer-Lernen

confusion matrix



Jamming Erkennung

- 4 Klassifizierungen pro Sekunde
- Single-Tone-Jammer (STJ), der um 13:11:09 Uhr ausgelöst wurde, wurde um 13:11:10 Uhr erkannt und wurde bis 13:11:19 Uhr kontinuierlich erkannt
- Genaue Erfassung der Jamming-Dauer
- Das Modell erkannte den Barrage- und Sub-Band-Jammer ebenso schnell

TABLE I
LOGS OF SINGLE-TONE JAMMING INSTANCES

Time	*Central Frequency (Hz)*	*Gain (dB)*	*Duration (s)*
13:11:09	3414100000	30	10
13:11:40	3429100000	26	13
13:12:11	3430100000	30	16
13:12:50	3417100000	28	12

Jamming Logs

TABLE II
TRUNCATED CLASSIFICATION LOGS

Time	*Class*	*Time*	*Class*	*Time*	*Class*
13:11:09	NS	13:13:52	NS	13:16:14	NS
13:11:10	STJ	13:13:53	SBJ	13:16:15	BJ
13:11:10	STJ	13:13:53	SBJ	13:16:15	BJ
13:11:10	STJ	13:13:53	SBJ	13:16:15	BJ
13:11:10	STJ	13:13:53	SBJ	13:16:15	BJ
...	...	...	...	...	...
13:11:19	STJ	13:13:54	SBJ	13:16:28	BJ
13:11:20	NS	13:13:55	NS	13:16:29	NS

Model Classification Logs

Agenda

Ein realer Fall

Ein realer Fall

Agenda

Fazit

Fazit/Take Home Messages

- KI kann helfen Angriffe zu erkennen
- SVM, Isolation Trees sind geeignet für Edge Geräte
- Komplexere ML Methoden sind eher für Geräte mit höherer Rechenleistung geeignet

- Die Bedrohungslage ist sehr ernst deshalb:
  - BSI Grundschutz
  - Mitarbeiter sensibilisieren
  - Patchen Patchen Patchen
  - Back-ups, Back-ups Back-ups

Danksagung

- Die vorgestellten Ergebnisse wurden teilweise durch folgende Projekte ermöglicht

KISS_KI

Mittelstand-Digital
Zentrum
Spreeland

Förderkennzeichen: 16KISK009, 01M023014C, 16KIS1204K

Danke für Ihre Aufmerksamkeit!

IHP GmbH – Leibniz-Institut für innovative Mikroelektronik
Im Technologiepark 25
15236 Frankfurt (Oder)
Tel.: +49 (0) 335 5625 350
E-Mail: langendoerfer@ihp-microelectronics.com