Cyber-Resilienz digitalisierter Energiesysteme

Einsatz von künstlicher Intelligenz in sicherheitskritischen CPS

Prof. Dr. Sebastian Lehnhoff

Das Stromsystem ist ein Cyber-Physisches System (CPS)

mit ganz besonderen Eigenschaften

1. **Wird zu den kritischen Infrastrukturen gezählt (KRITIS)**
   > Unverzichtbare Lebensader moderner Gesellschaften
2. **Kontinentübergreifende Größe**
   > Von Nordafrika bis Skandinavien, von Irland bis Asien
3. **Ausbreitung von Phänomenen und deren Dynamik**
   > Instantane Ausbreitungsgeschwindigkeit von Instabilitäten
4. **Allgegenwärtige Zielkonflikte**
   > Monetäre, technische, (nationale/internationale) politische Interessen
5. **Im schnellen und grundlegenden Wandel begriffen...**

Drei große Trends beeinflussen die Stabilität der KRITIS Energie

1. **Umbau des Energiesystems, u. a.**
   - > viele kleinere Anlagen, in der Gesamtheit systemkritisch
   - > Wettbewerb und neue Geschäftsmodelle
   - > Vernetzung durch Digitalisierung
2. **Digitalisierungstrends, u. a.**
   - > Internet of Things (IoT): mehrere Mrd. Geräte im Internet und mit unserem Stromnetz verbunden (Fernseher, Babyphon, Aquarien usw.)
   - > Smart Services, Cloud, Outsourcing, Künstliche Intelligenz, Big Data,…
3. **Neue Bedrohungslage für Cyberattacken, u. a.**
   - > staatlich unterstützte Cyberattacken („Grey War“) mit
   - > ausgefeilten Werkzeuge (teilweise staatlich hergestellt)
   - > staatlicher Wunsch nach „Hintertüren“ (BMI, Homeland Security,…)

Folgerungen

ESYS-AG "Zuverlässigkeit digitalisierter Energiesysteme"

1. Digitalisierungstrends und -herausforderungen aus anderen Bereichen werden zwangsläufig auf die Energieversorgung überschwappen.
2. Die Digitalisierung der Energieversorgung führt zu einer neuen Bedrohungslage mit ungeahntem gesellschaftlichem Schadenspotential.
3. Die Digitalisierung der Energieversorgung ist zugleich aber notwendig für die Energiewende und bietet große Chancen – auch für die Sicherheit.

Und: *Wir verstehen den Zusammenhang zwischen Digitalisierung und KRITIS noch nicht wirklich!*

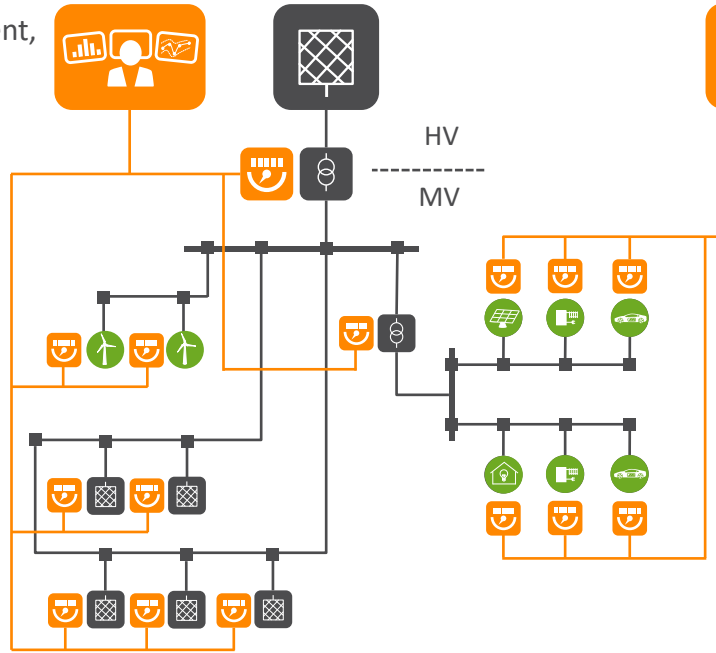
Leopoldina Nationale Akademie der Wissenschaften
acatech DEUTSCHE AKADEMIE DER TECHNIKWISSENSCHAFTEN
UNION DER DEUTSCHEN AKADEMIEN DER WISSENSCHAFTEN

Energiesysteme der Zukunft (ESYS II)

Energiesysteme sind komplexe cyber-physische Systeme

Vielfältige Aufgaben in heterogenen, verteilten (Teil-) Systemen

- Prognose von Netzzuständen,
- optimiertes Blindleistungsmanagement,
- Erkennen von Anomalien in Strom- und Kommunikationsnetz.

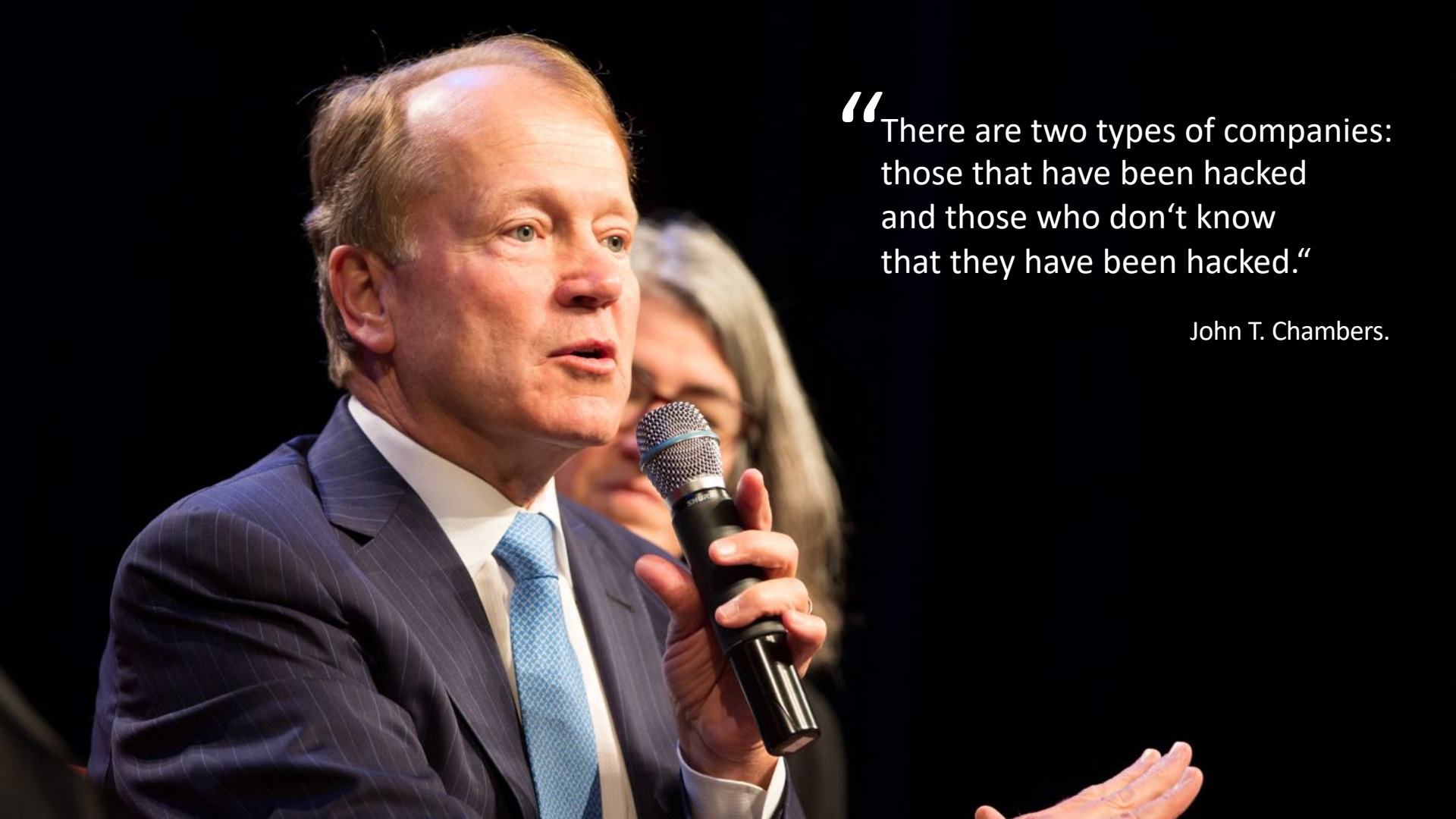


- Algorithmischer Energiehandel,
- Marktintegration Erneuerbarer Energien.

- Virtuelle Kraftwerke,
- multi-modale Optimierung,
- Sektorkopplung.

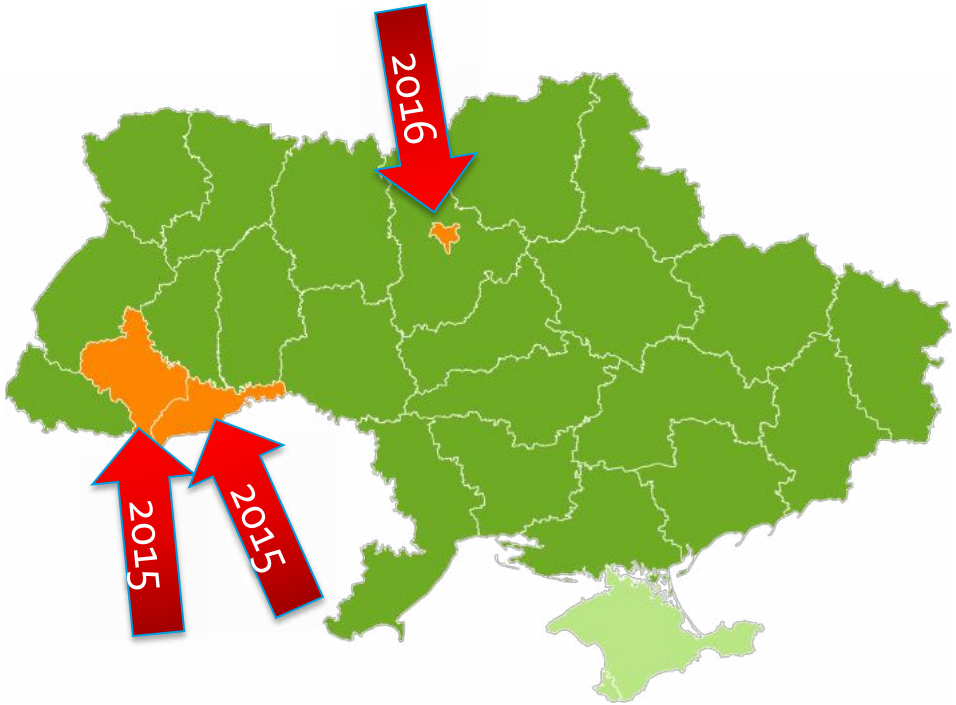
- Überwachung der Betriebszustände,
- Automatisierung gelbe Ampelphase,
- dezentrale Systemdienstleistungen.

- Last- und Flexibilitätsmanagement,
- Einbindung von Endkunden,
- Smart Metering.


“There are two types of companies:
those that have been hacked
and those who don‘t know
that they have been hacked.“
John T. Chambers.

Das trifft auch auf Energiesysteme zu

Hackerangriff auf das Stromnetz in der Ukraine, 2015

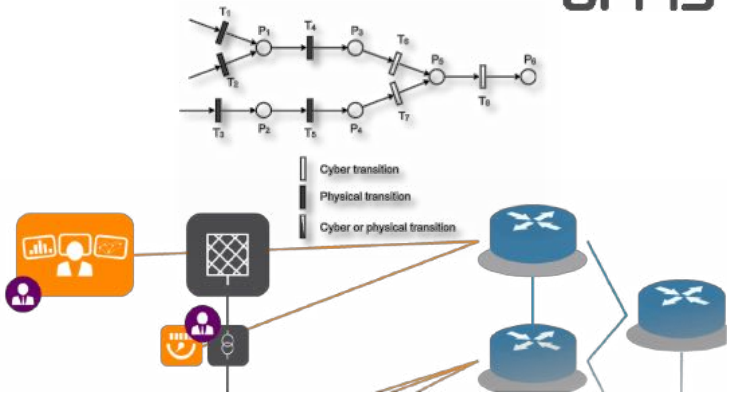


23.12.2015

- Blackout in Ukraine durch **Hackerangriff**
- **3 Stromversorger** betroffen
- **Hoher Automatisierungsgrad** der Verteilnetze
- Operativer Eingriff in die **Netzleittechnik** und **Abkopplung mehrerer Umspannstationen** vom Netz
- **Mehrere Monate** Vorbereitung

Statische Modellierung von CPES ist nicht ausreichend

State of the Art



Fallbasierte Modellierung

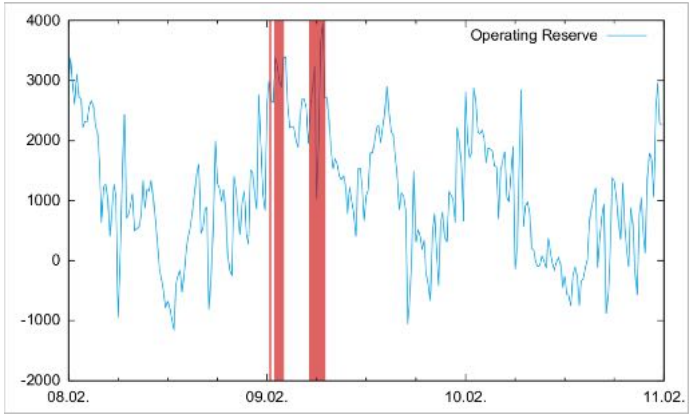
Hierarchische Herangehensweise

Auch bei komplexen CPS eindeutiger Fokus, z.B.

- > **IKT**: „Ressourcen für Kommunikation?“
- > **Energie**: „Ressourcen für Energiebereitstellung?“

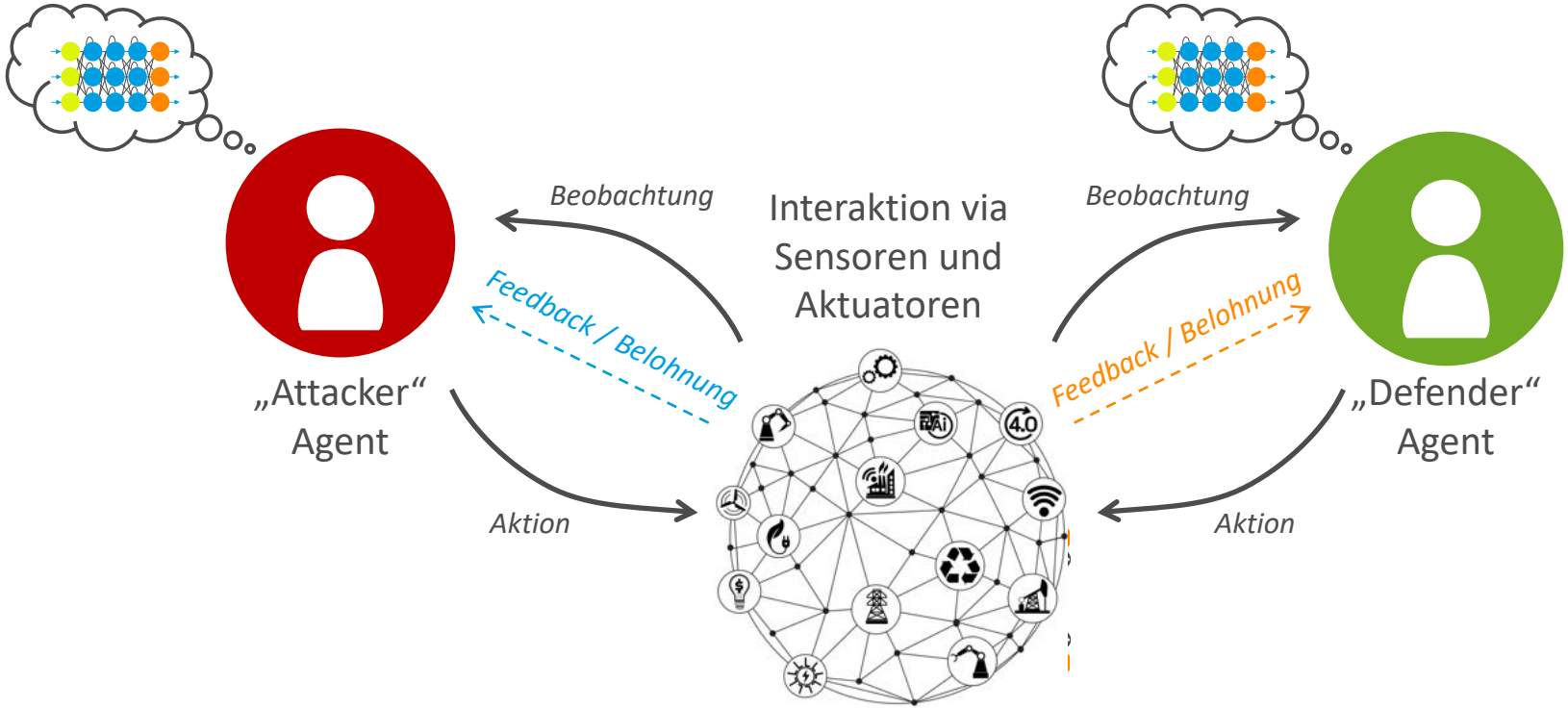
Marktperspektive selten berücksichtigt

- > Markteinfluss eindeutig gegeben
- > 09. Feb 2012: Beinahe Blackout in Süddeutschland wegen irregulärer Nutzung von Ausgleichsenergie (> 80% der verfügbaren Regelleistung abgerufen)



Das Konzept des Adversarial Resilience Learning

Konkurrierende Agenten lernen durch Interaktion auf einer geteilten Umgebung



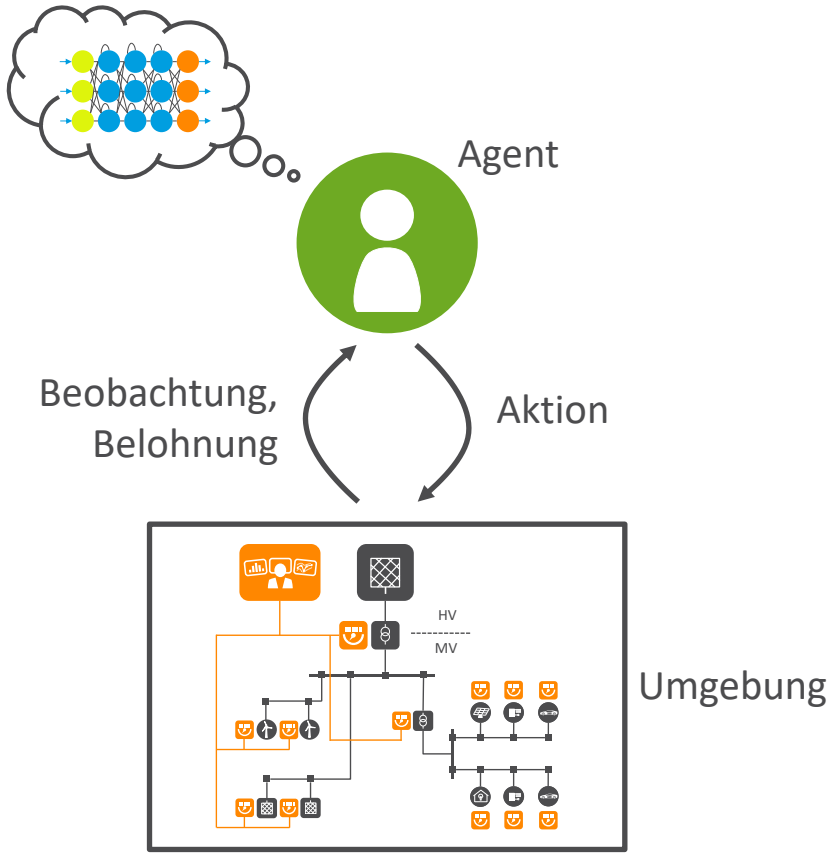
Definition geeigneter Belohnungsfunktionen

Am Beispiel der Systemperformanz eines Smart Grids

Training auf Basis einer geeigneten Belohnungsfunktion, die Indikatoren der Performanz eines CPS umfasst:

- Knotenspannungen,
- Leitungsströme,
- Blindleistungsbilanz an Übergabepunkten zu höheren Netzebenen,
- Energie-/Regelleistungspreise
- ...

? **Offene Forschungsfrage:** Metrik für Resilienz eines Smart Grids?



Anwendung: Analyse und Training resilienter Systeme

Die Zuschnitte von ARL

Analyse – nur Angreifer

- Testlabor für resiliente Systeme
- Angreifer exploriert Schwachstellen
- „Eroberung“ des Systems
- Angriffsvektoren und -log als Analysegrundlage

ARL & Ethik

- ARL als „Angriffswaffe“?
- **Lizenz** als Lösungsansatz?
- **Robotergesetze** per Transfer Learning inhärent machen?

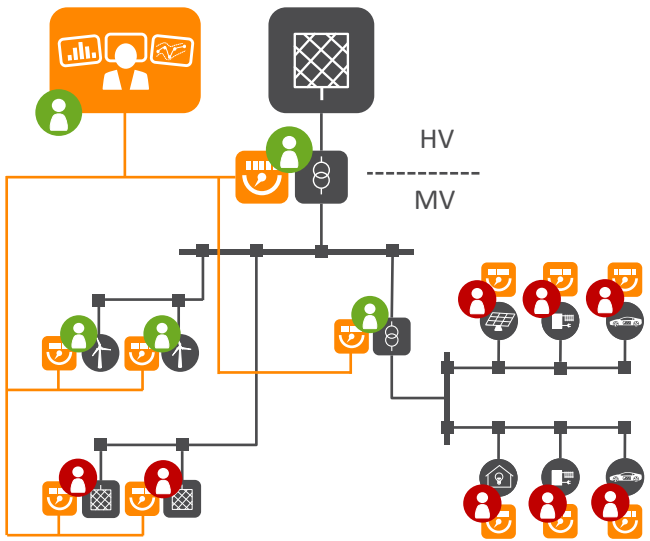
Training – Angreifer und Verteidiger

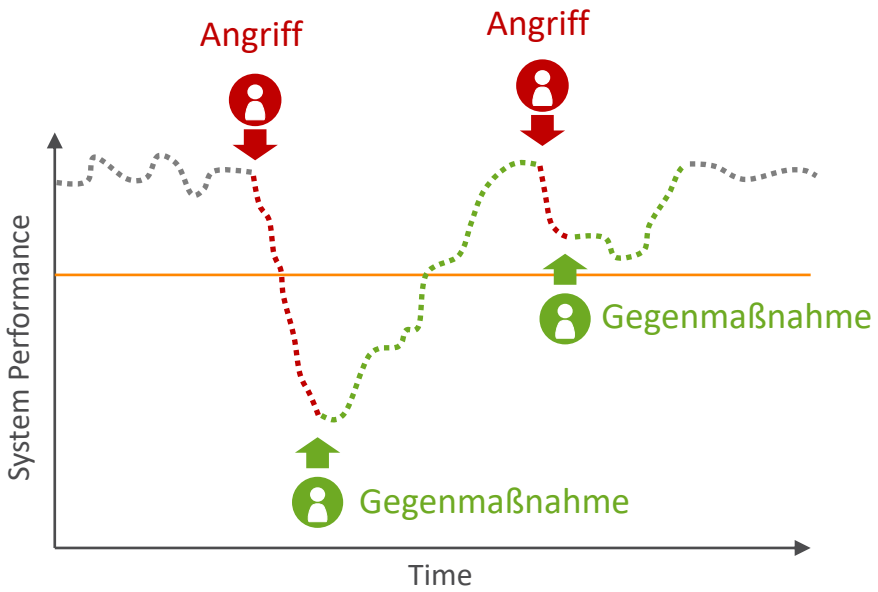
- KI-Betriebsführung
- Resilienz des Gesamtsystems
- Angreifer trainiert Verteidiger
- Angriffe: nicht nur böswillig, auch natürliche Umweltfaktoren
  - Prognoseabweichungen
  - Schäden durch Unfälle o.ä.

Demo: Angriff aufs Stromnetz

Vereitelung von (Teil)Systemübernahme als nachrangiges Problem

Cyber Resilience Lab





Attacker AI

Defender AI

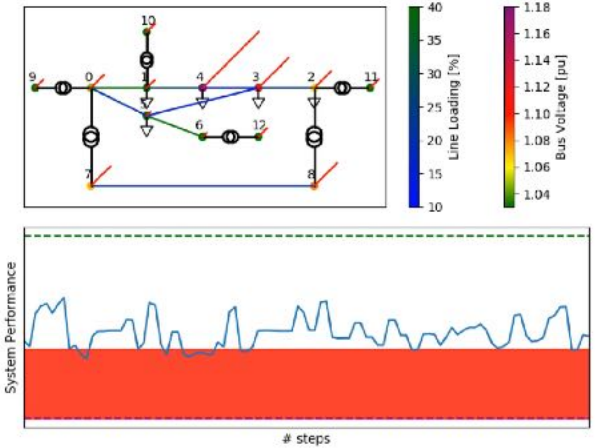
Demonstrator für Adversarial Resilience Learning

KI-basierte Analyse der Resilienz von Smart Grids

```
Attacker measures (2000/3000)
-- grid.load0.change(scaling=0.76)
-- grid.load1.change(scaling=0.92)
-- grid.load1.change(scaling=0.17)
-- grid.load1.change(scaling=0.41)
Learning: Attacker ...
```

Attacker Score:

0




```
Defender measures (2000/3000)
-- grid.gen0.change(scaling=0.37)
-- grid.gen1.change(scaling=0.1)
-- grid.trafo0.change(tp_pos=7.0)
-- grid.trafo1.change(tp_pos=-9.0)
-- grid.trafo2.change(tp_pos=5.0)
-- grid.trafo3.change(tp_pos=-5.0)
-- grid.trafo4.change(tp_pos=8.0)
-- grid.trafo5.change(tp_pos=-1.0)
Learning: Defender ...
```

Defender Score:

0

Fazit und Ausblick

ARL erlaubt das Lernen von Schwachstellen und Zusammenhängen

> Auch bei technisch, regulatorisch konformem Verhalten (EnWG, GridCodes, TAB etc.)

Entwicklung von Verteidigungs(!)strategien

> Ethisches Dilemma?

„Attacker-Defender-Games“

> Impact-Analyse in „Anomalie-sensitiver State Estimation“

> Risikomodelle und Investitionsentscheidungen (zur Erreichung eines Gleichgewichtszustandes)

> Untersuchung von Asymmetrien („Rigging the Game“)

Smart Grid Cyber Resilience Lab

Gefördert durch:
Bundesministerium für Wirtschaft und Energie
aufgrund eines Beschlusses des Deutschen Bundestages