Three Major Trends

...that influence the stability of KRITIS Energy

1. **Conversion of the energy system, e.g.**
   > Many smaller systems, system-critical as a whole
   > Competition and new business models
   > Interconnectivity, flexibilization through digitalization

2. **Digitization trends, including**
   > Internet of Things (IoT): several billion devices on the Internet and connected to our power grid (televisions, baby monitors, alexa, etc.)
   > Smart Services, Cloud, Outsourcing, Artificial Intelligence, Big Data,...

3. **Susceptibility to new effects, e.g.**
   > Occurrence of "classic" IT challenges (errors, update management, interactions)
   > Sophisticated cyber-attacks (partly supported by states)
Energy Systems are Complex Cyber-Physical Systems
Diverse tasks in heterogeneous, distributed (sub)systems under different responsibilities

- Forecast of network conditions,
- Optimized reactive power management,
- Detection of anomalies in power and communication networks.

- Monitoring of the operating states,
- Automation yellow traffic light phase,
- Decentralised system services.

- Load and flexibility management,
- Integration of end customers,
- Smart metering.

- Virtual power plants,
- Multi-modal optimization,
- Sector coupling.

- Algorithmic energy trading,
- Market integration of renewable energies.

- Algorithmic energy trading,
- Market integration of renewable energies.

A wide range of entry-points into a safety-critical infrastructure...
“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.
This also applies to „our“ Energy Systems
Cyber-attacks on the power system in the Ukraine, 23.12.2015 (and subsequently in 2016)

- Blackout in Ukraine due to hacker attack
- 3 Power utilities affected
- Operative manipulation of the automation systems and decoupling of several transformer stations from the network
- Several months of preparation

> Power Systems are high-value targets: how to reliably detect vulnerabilities?
Vulnerabilities
...to critical dependencies and cyber-attacks

Critical attack vectors in energy systems (non exhaustive)
> Reconnaissance, data theft
> IT/OT-hacking: remote access and control
> Data-spoofing: bad data injection, data manipulation, excitation of dynamic instabilities/sliding modes

Existing monitoring systems
> Intrusion/anomaly detection \(\rightarrow\) abnormal network traffic
> State estimation \(\rightarrow\) measurement outliers (statistical), varying accuracy of measurements

Not sufficient for detecting critical situations in digitalized energy systems!
State Assessment based on Trust Facets

<table>
<thead>
<tr>
<th>Trust Facets</th>
<th>Measurement Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Functional Correctness</td>
<td></td>
</tr>
<tr>
<td>II. Safety</td>
<td></td>
</tr>
<tr>
<td>III. Security</td>
<td></td>
</tr>
<tr>
<td>IV. Reliability</td>
<td></td>
</tr>
<tr>
<td>V. Credibility</td>
<td></td>
</tr>
<tr>
<td>VI. Usability, Understandability</td>
<td></td>
</tr>
</tbody>
</table>

All conceivable attack vectors manifest themselves in a combination of (violated) trust facets

What to do with this multivariate assessment?

- E.g. substitute measurements with historical/simulated values?
- Do nothing?

What is the worst that could happen?
The Concept of Adversarial Resilience Learning
Competing agents learn by interacting on a shared environment

Shared Environment
(Digital Twin of a CPS)

Interaction via Sensors and Actuators

“Attacker” Agent

“Defender” Agent

Observation

Feedback / Reward

Action

Observation

Feedback / Reward

Action
Use Case: Resilient Systems Analysis and Training

Variations of ARL

**Analysis – only Attacker**
- Test laboratory for resilient systems
- Attacker explores vulnerabilities
- "Conquest" of the system
- Attack vectors/results as a basis for analysis

**Training – Attacker and Defender**
- AI for automated operation
- Resilience strategies of the overall system
- Attacker trains defender
- Attacks: not only malicious, but also natural environmental factors
  - forecast deviations
  - Damage caused by accidents etc.

**What about the obvious ethical dilemma?**
- ARL as "assault weapon"?
- License as a solution?
- Making “laws of robotics” inherent by transfer learning?
Demo: Attack on a Power System
Prevention of (sub-)system takeover as a secondary problem

Attacker AI  Defender AI
Conclusion and Outlook
Digitalization is indispensable for flexible energy systems (and highly vulnerable at the same time)

**Traditional means/methods have been proven to miss:**
- Vulnerabilities to interdependent/dynamic failures
- Specialized/targeted attacks

**Multivariate impact analysis necessary**
- Basis for (automated) decision-making during operation ("always compromised")
- *Risk-based investments in countermeasures?*

**ARL as an AI-based game-theoretic approach to vulnerability testing (CPS modelling)**
- Equilibria more relevant (and achievable) than "absolute" safety

**There is no way back from digitalized energy systems!**
- Most promising answer against highly specialized/targeted attacks is *Operational Flexibility (on-line change of system characteristics) for Cyber-Resilience!*