Adversarial Resilience Learning

Das Konzept des *Adversarial Resilience Learning* (ARL) definiert zwei Klassen von gegeneinander agierenden Agenten, die auf ein gemeinsames, komplexes *Cyber-physikalisches System* (CPS) – hier das Stromnetz – einwirken. Abstraktes Ziel der *Attacker Agents* ist es, negativ im Sinne einer Systemperformanz auf das CPS einzuwirken. Dem entgegengesetzt ist es Ziel der *Defender Agents,* die Systemperformanz des CPS in einem gewünschten Bereich zu halten bzw. das CPS aus ungünstigen Systemzuständen wieder in einen wünschenswerten Betriebszustand zu überführen. Damit können zum einen im Rahmen von *Resilienzanalysen* für bestehende CPS systematisch Schwachstellen und Risiken ermittelt werden, indem ein oder mehrere *Attacker* auf das System einwirken und versuchen, einen möglichst ungünstigen Systemzustand herzustellen bzw. einen möglichst großen Schaden anzurichten. Zum anderen können im Rahmen eines *Resilienztrainings* KI-basierte *Defender* entwickelt und trainiert werden, die die Funktion des CPS trotz der Angriffe und destabilisierenden Maßnahmen der *Attacker* aufrechterhalten. Durch fortlaufendes Training entsteht eine Software, die die Resilienz des CPS maßgeblich unterstützt.