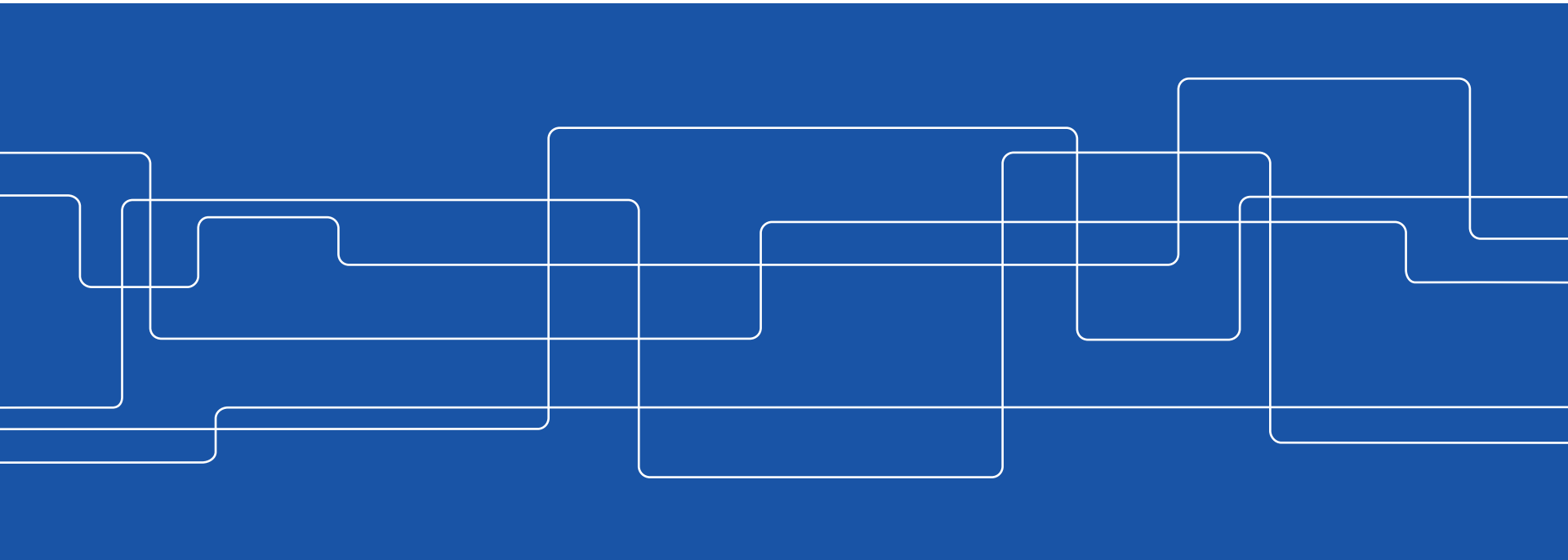
KTH ROYAL INSTITUTE
OF TECHNOLOGY

Physics-Based Attack Detection and Countermeasures in Control Systems

Henrik Sandberg
Department of Automatic Control
KTH, Stockholm, Sweden



In Collaboration With…

KTH and CERCES:
György Dán, Ragnar Thobaben, Mads Dam,
Kaveh Paridari, Jezdimir Milošević,
David Umsonst, Karl Henrik Johansson







Delft University of Technology:
André M.H. Teixeira

University of Texas at Dallas:
Alvaro A. Cárdenas, and co-workers

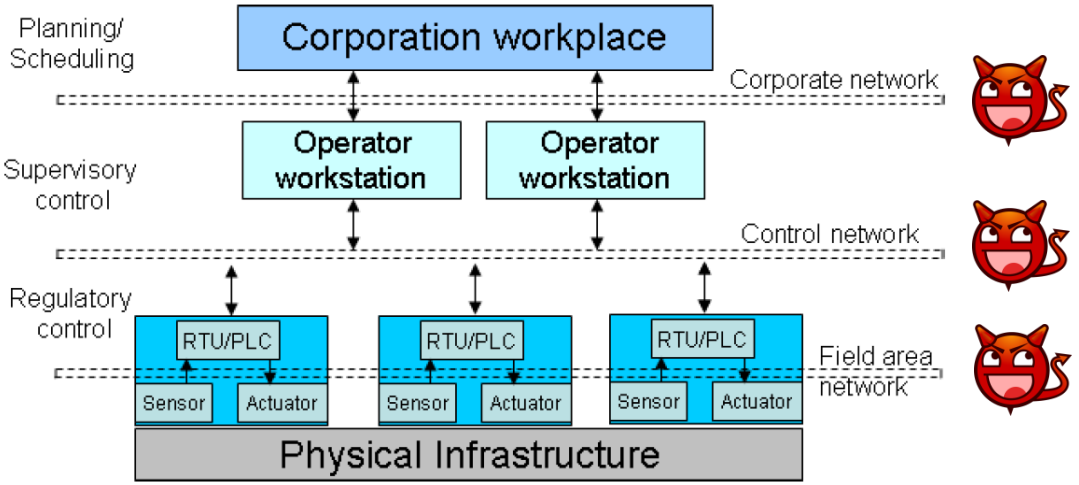
SPARKS (EU FP7):
AIT, UTRC, and EMC Corporation

SMART GRID PROTECTION AGAINST CYBER ATTACKS

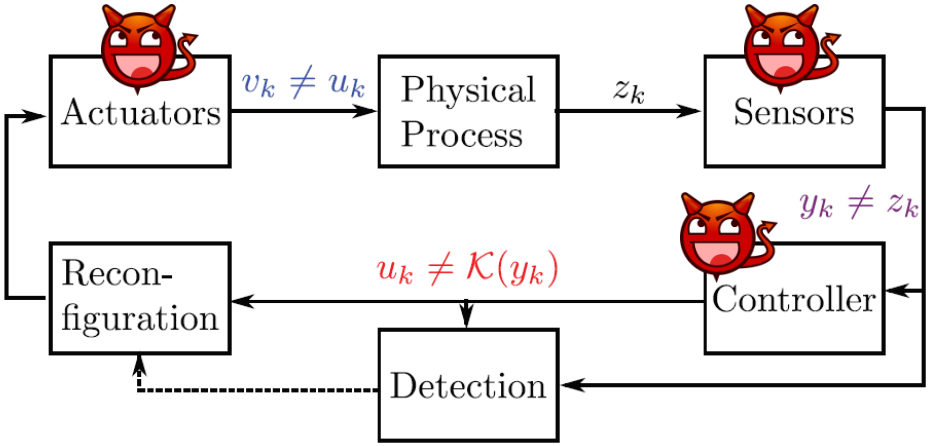
[Cardenas *et al*., Hotsec '08]
[Urbina *et al*., CCS '16]

Industrial Control System (ICS) under Attack

IT perspective:



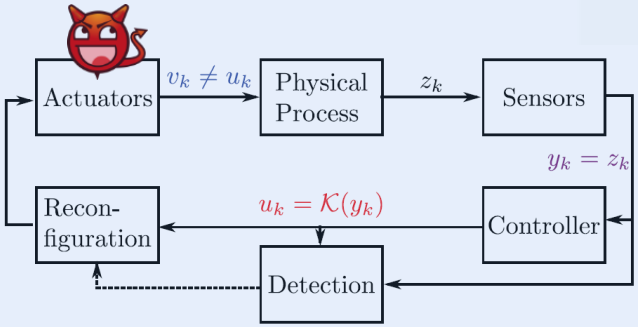
Control perspective:



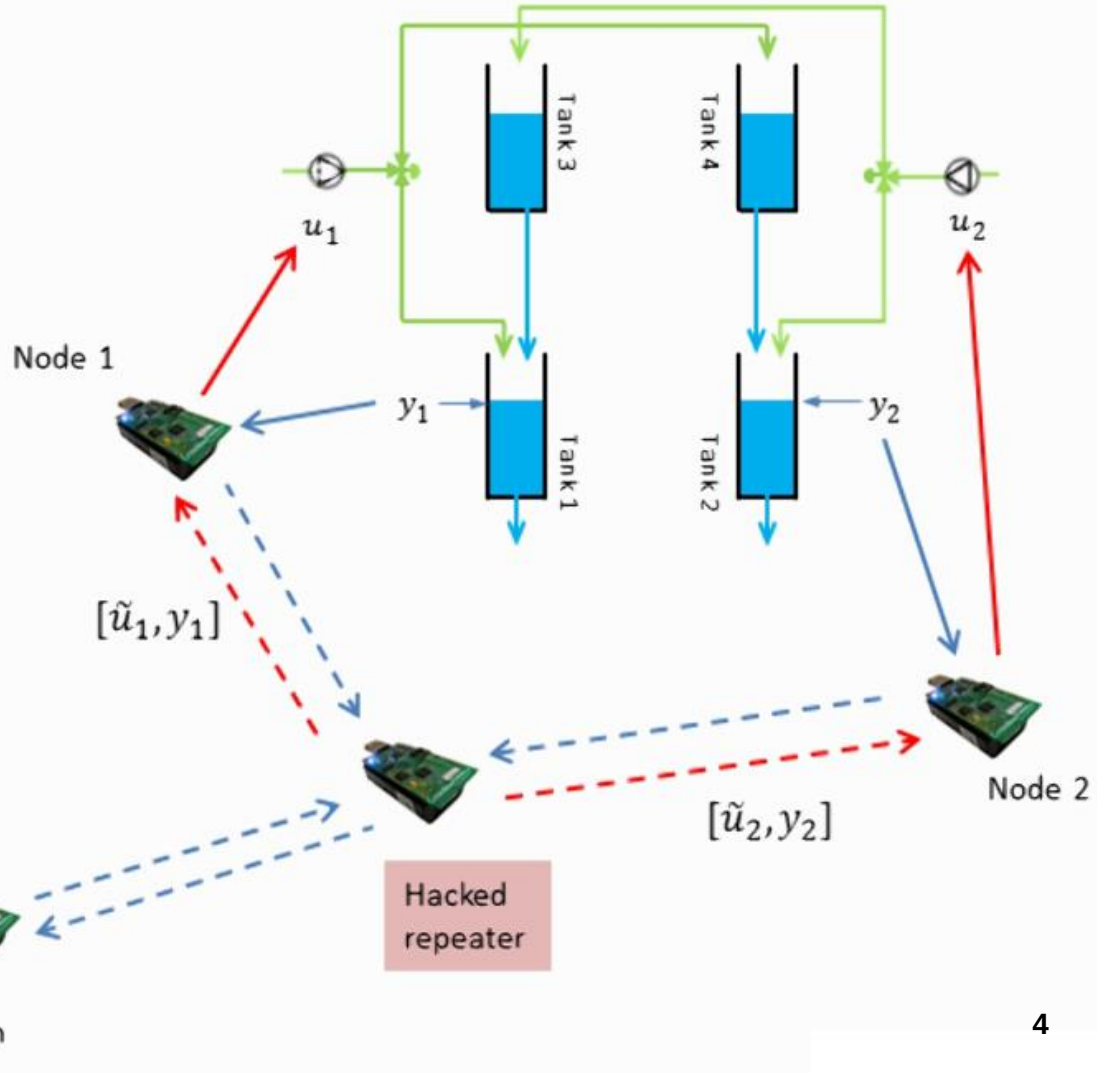
[Teixeira *et al*., HiCoNS '12]

Example: Stealthy Water Tank Attack

2 hacked actuators (u_1 and u_2)
2 healthy sensors (y_1 and y_2)

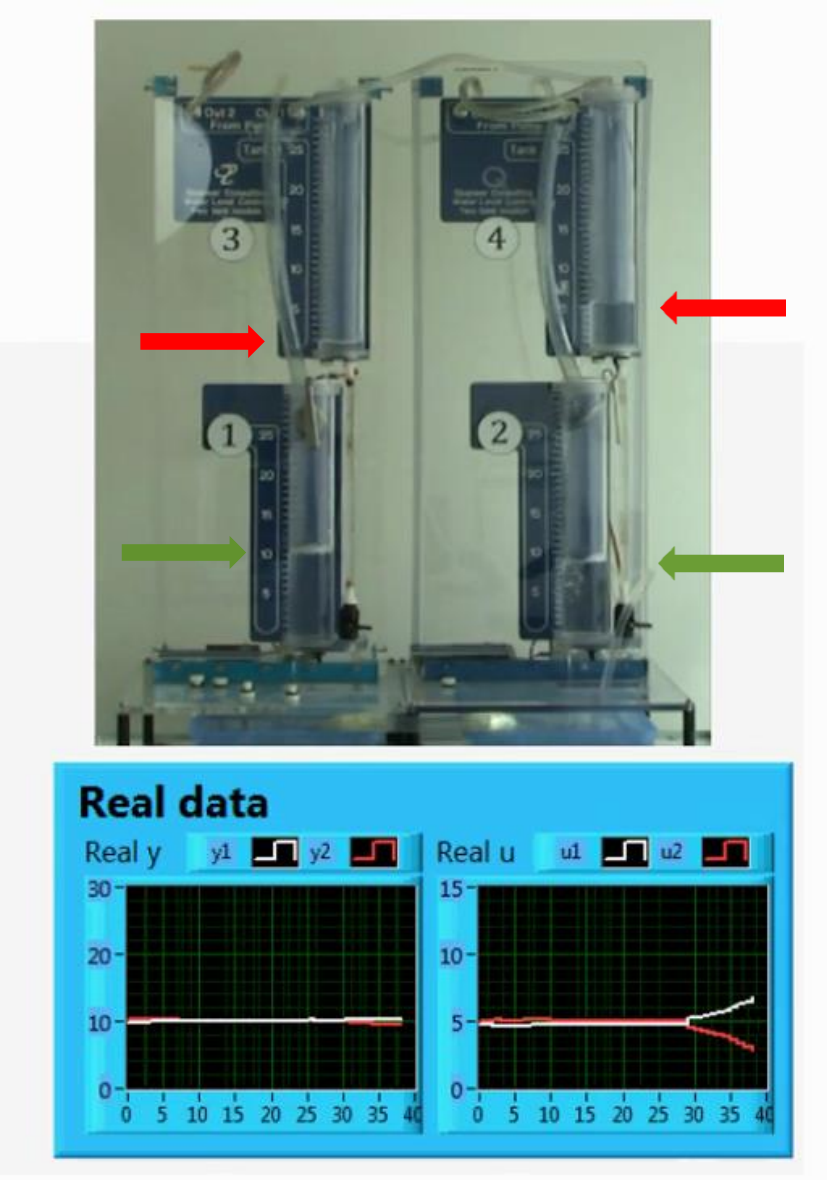


Can the controller/detector always detect the attack?



[Teixeira *et al*., HiCoNS '12]

Example: Stealthy Water Tank Attack [Movie]



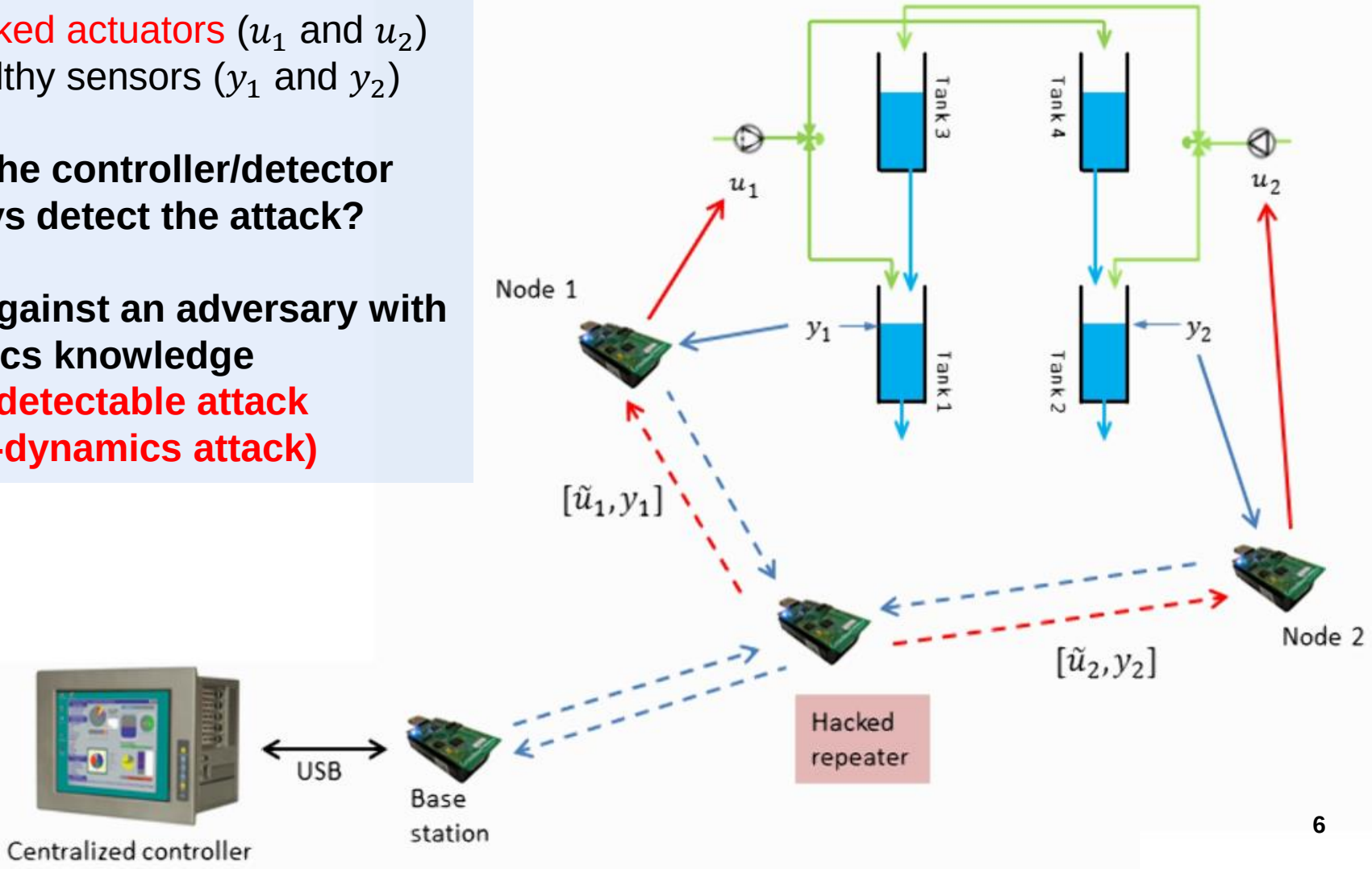
[Teixeira *et al*., HiCoNS '12]

Example: Stealthy Water Tank Attack

2 hacked actuators (u_1 and u_2)
2 healthy sensors (y_1 and y_2)

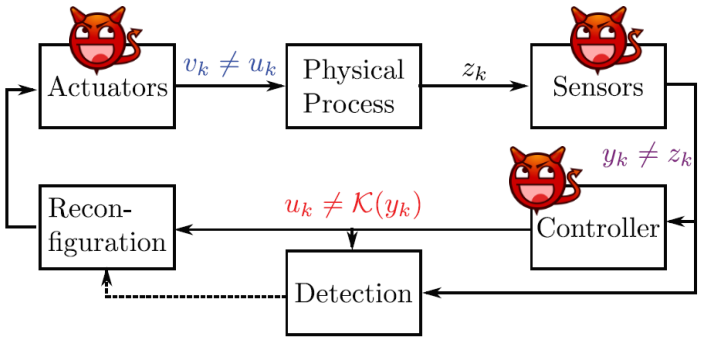
Can the controller/detector always detect the attack?

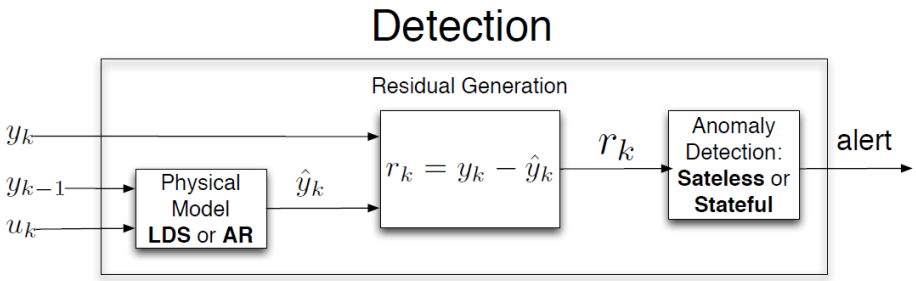
Not against an adversary with physics knowledge
⇒ Undetectable attack (zero-dynamics attack)



[Urbina *et al.*, CCS '16]

Physics-Based Anomaly Detection



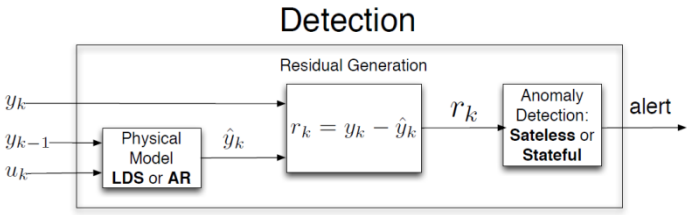


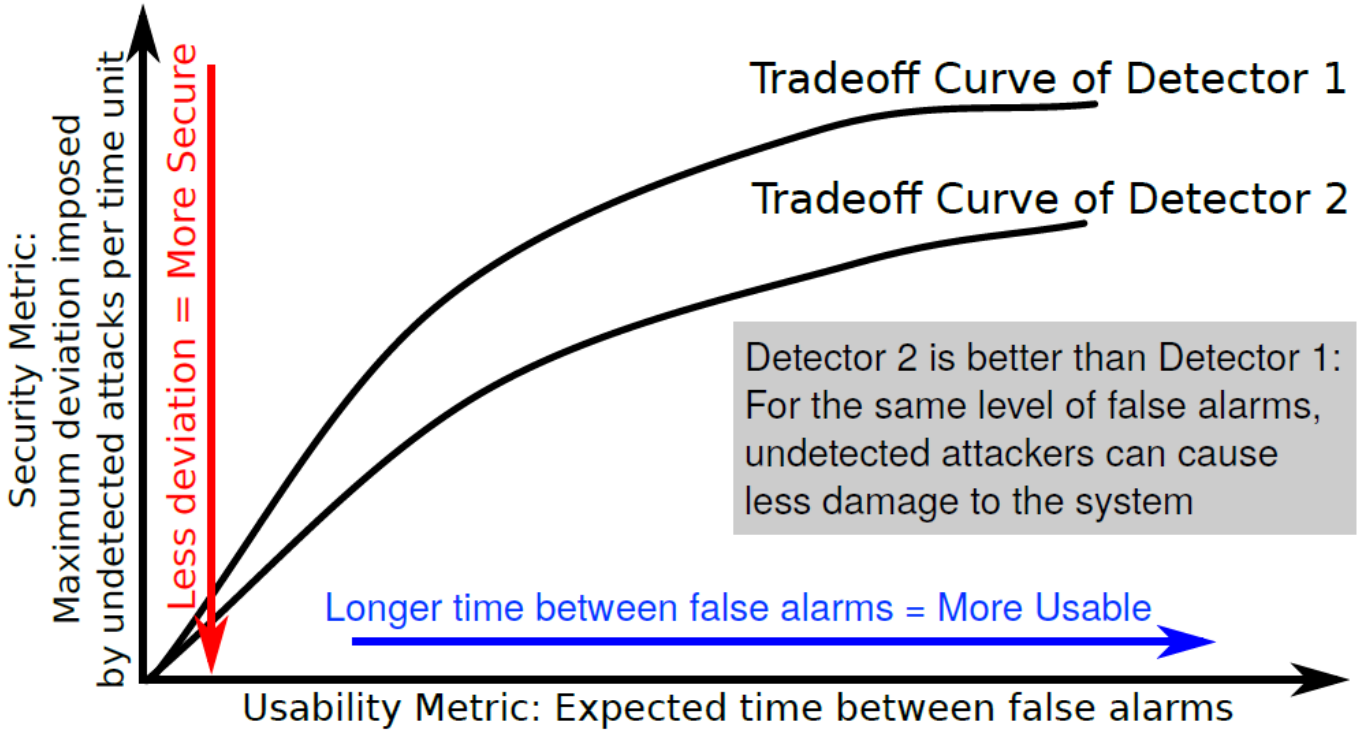
- Physics-based anomaly detectors work for
  - Randomly failing components [safety]; and
  - Physics-unaware adversaries [security]
- **But** example illustrates sensitivity to adversaries with
  - *Physical process knowledge*; and ability to stage *coordinated* (time & space) data corruption [security]

- Quantify performance of and compare different detectors?

New Performance Metric for ICS Anomaly Detection

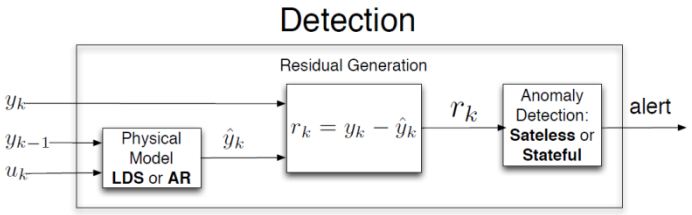
[Urbina *et al.*, CCS '16]

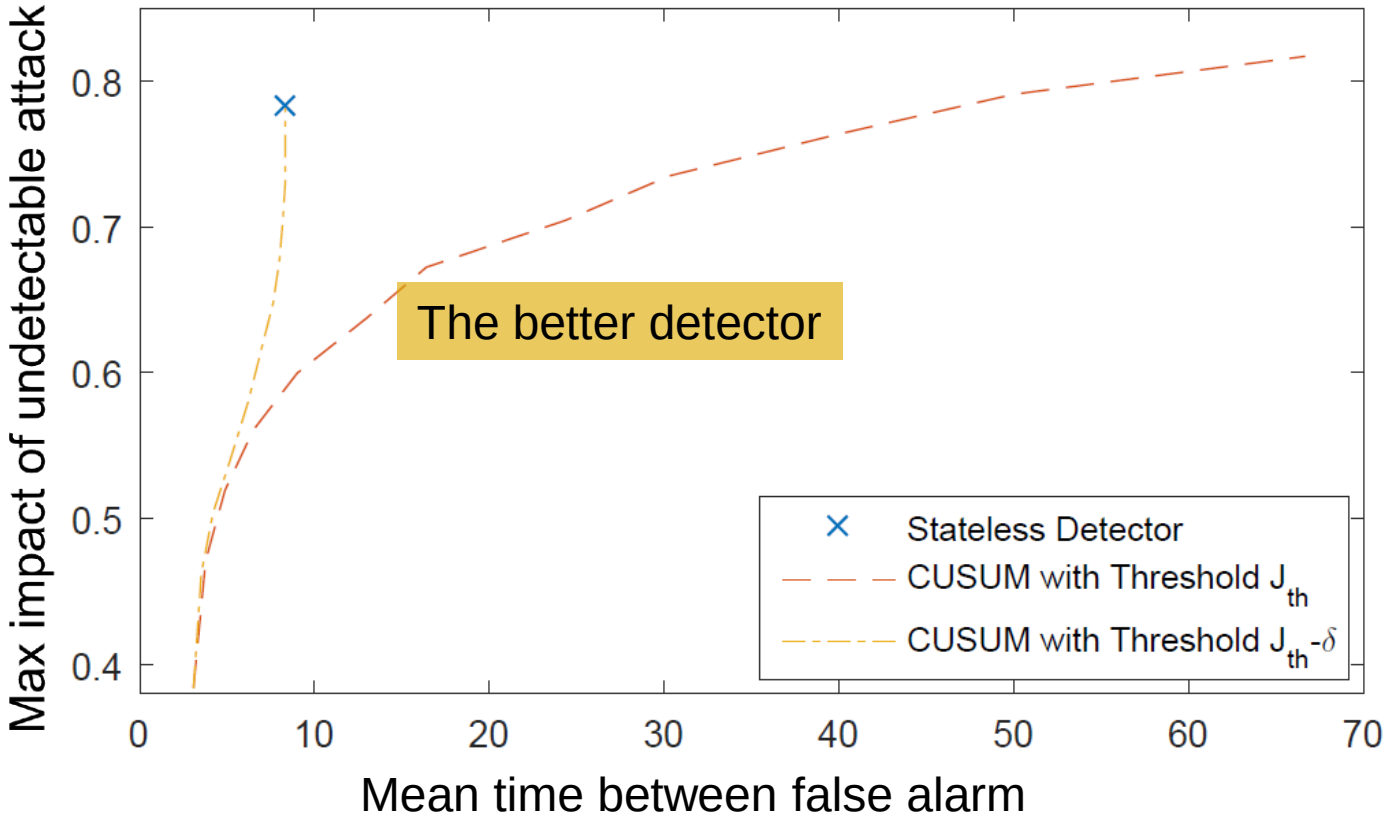




Power System Example

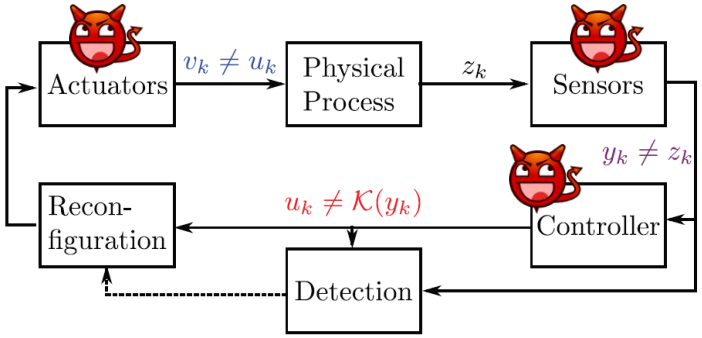
[Umsonst *et al*., submitted ‘16]

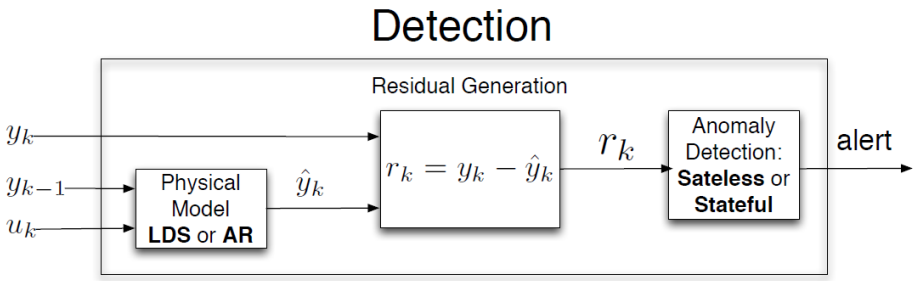




(No attack and no component failure, caused by “normal” process and sensor noise)

Physics-Based Attack Detection and Countermeasures in Control Systems





What can we do in real time about the attacks and faults *we can detect* using the anomaly detector?

I.e., what about the countermeasures (=reconfiguration)?

Example next…

A Test-bed and Case Study: NIMBUS Microgrid, Cork, Ireland

SMART GRID PROTECTION AGAINST CYBER ATTACKS

Electrical components

10kW wind turbine

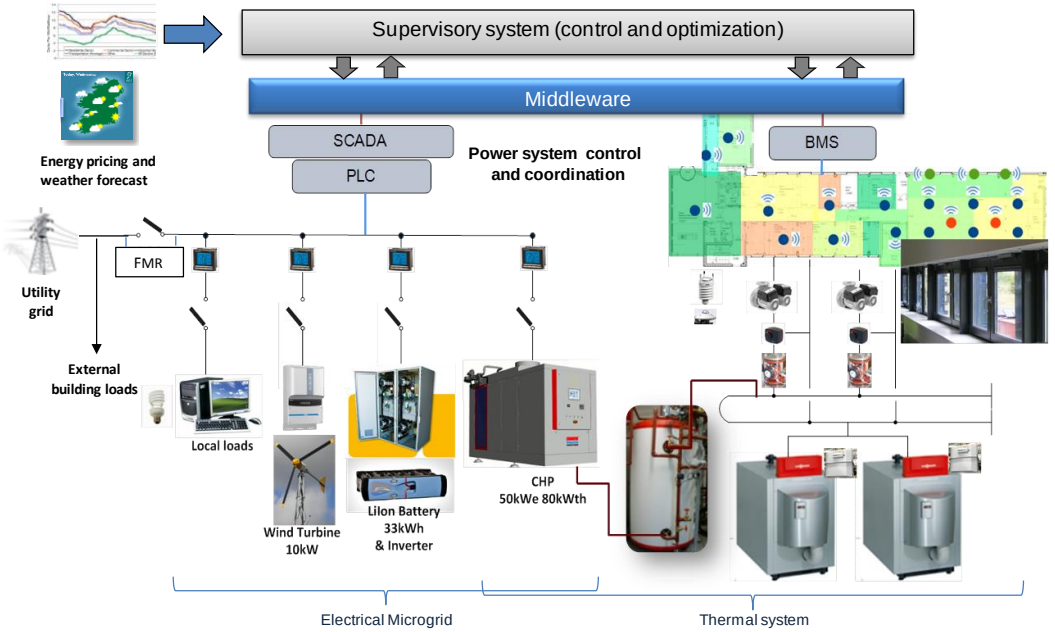
35kWh (85kW peak) Li-Ion battery

50kW electrical/82kW thermal combined heat and power unit (CHP) and

Feeder management relay to manage the point of coupling between the microgrid and the rest of the building, and a set of local loads.

Battery and wind turbine interfaced through power electronics converters

CHP with synchronous machine



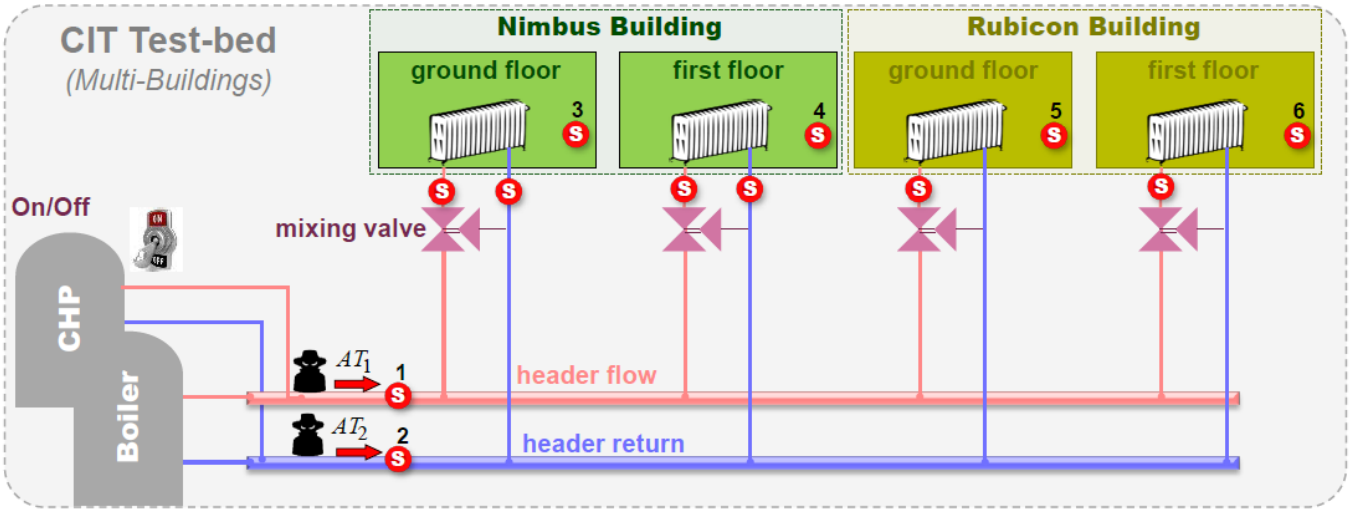
IT System

Interlinked Building Management System and Microgrid SCADA

Three-layer control systems

UTRC Middleware

Attack Scenario



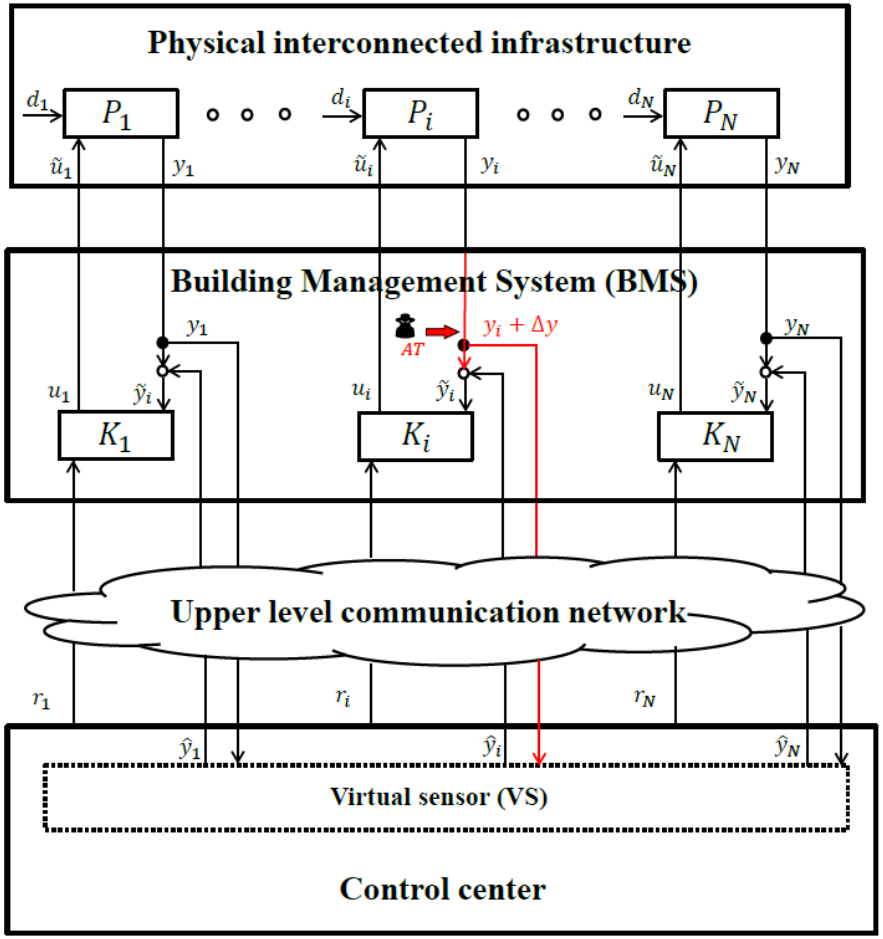
Adversary: Infect some field devices with malware (á la Stuxnet) corrupting measurements sent to PLCs (Here: AT_1 and AT_2)

Defender: Access to remote correlated measurements and a physical model (here temp. measurements and modeling by system identification)

In collaboration with UTRC and EMC Corporation (Ireland)

[Paridari *et al.*, ICCPS '16]

Resilient Monitoring and Control



1. Anomaly detector in control center detects attacked measurement $y_i + \Delta y$

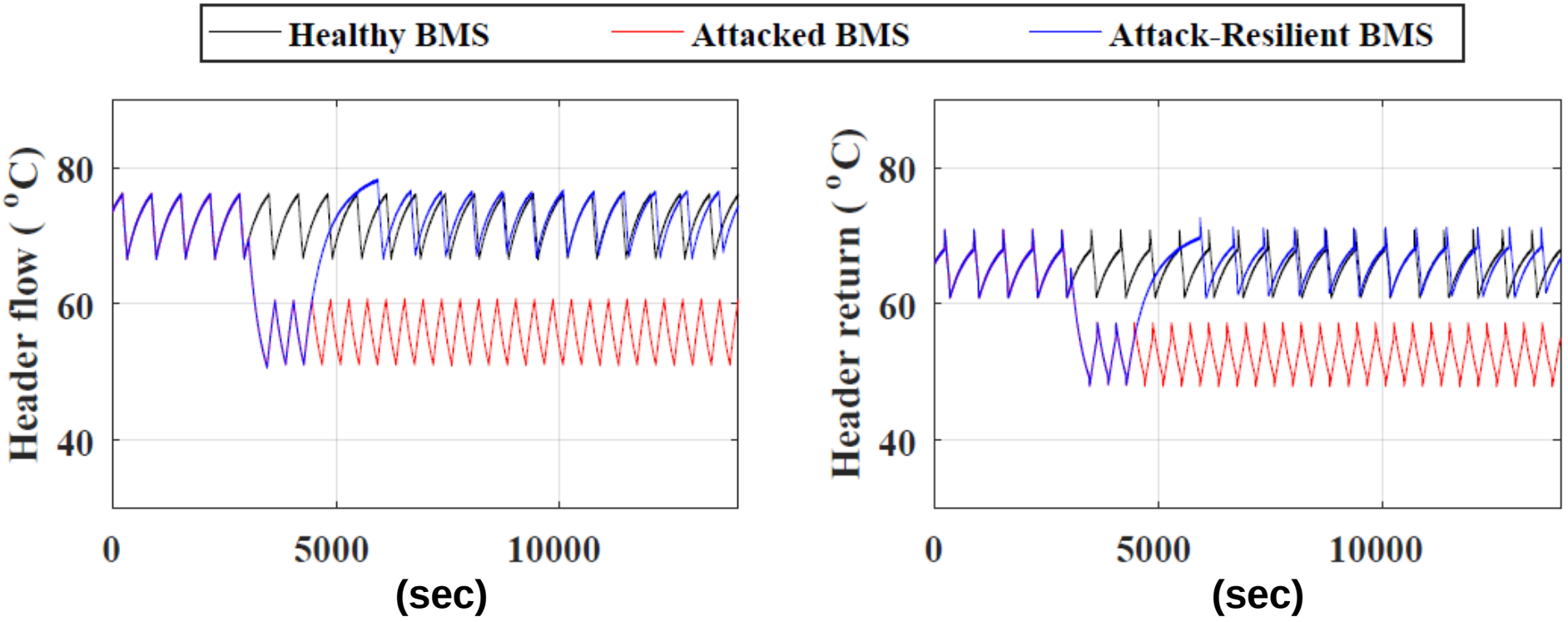
2. Optimal physics-based prediction $\hat{y}_i$ from **un-attacked** measurements $y_1, .., y_N$ (Virtual sensor)

3. Feed $\hat{y}_i$ back to PLCs

[Paridari *et al.*, ICCPS '16]

Verification: Control Performance

1400 sec delay in anomaly detector (“attacker free time”):

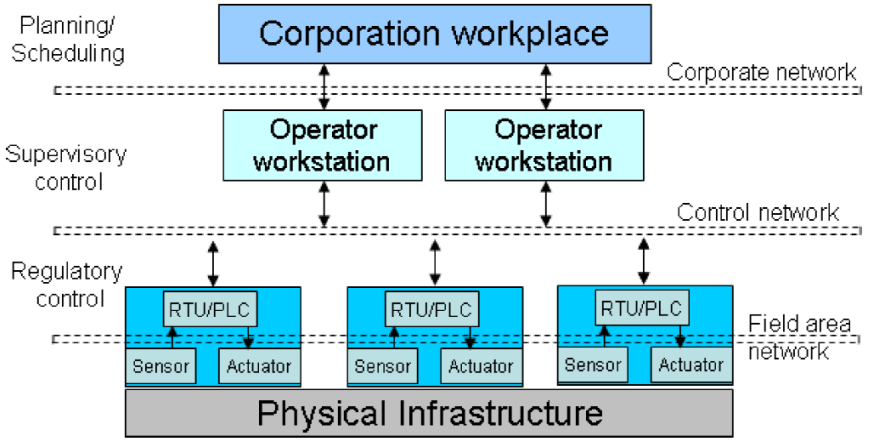


[Paridari *et al.*, ICCPS ‘16]

Summary

- Possibilities with physics-based anomaly detectors:
  - Randomly failing components [safety]: **OK**
  - Physics-unaware adversaries [security]: **OK**
  - Adversaries with *physics knowledge* and ability to stage *coordinated* (time & space) data corruption [security]: **not always OK (example in movie)**
- New metric to evaluate anomaly detectors for ICS. Tools under development
- Fault- and attack-tolerant (resilient) controller example

CERCES – Center for Resilient Critical Infrastructures



[Area 4]
[Area 3]
[Area 2]
[Area 1]

Figure 1. Architecture of control systems.

- Area 1: Embedded Software Platforms (M. Dam)
- Area 2: Wireless Communication (R. Thobaben)
- Area 3: Communication and Computation Infrastructure (G. Dán)
- Area 4: Resilient Control of Cyber-Physical Systems (H. Sandberg)

Thank You!

- CERCES: www.ees.kth.se/cerces

- SPARKS: project-sparks.eu/

- Henrik Sandberg: people.kth.se/~hsan/