The Threat of AI-Driven Smart Malware

The Case of Availability Attacks on Computing Systems through Alteration of Environmental Control

ILLINOIS
Electrical & Computer Engineering
COLLEGE OF ENGINEERING

Zbigniew Kalbarczyk
kalbarcz@illinois.edu

Collaborators: Key-Whan Chung, Ravi Iyer

Advanced Targeted Attacks => AI-Driven Smart Malware

- *Malicious actors:*
  - Actively learn a target entity’s infrastructure and normal behavior/operations and use this knowledge to devise an attack strategy
  - Dedicate an effort to maintain anonymity and stay beyond the radar of security monitoring system
  - Highly sophisticated - expertise in a target system/infrastructure
  - Patient in preparing and executing attack strategy, i.e., malicious activities may span a long time
  - Agile to work around victim’s defenses

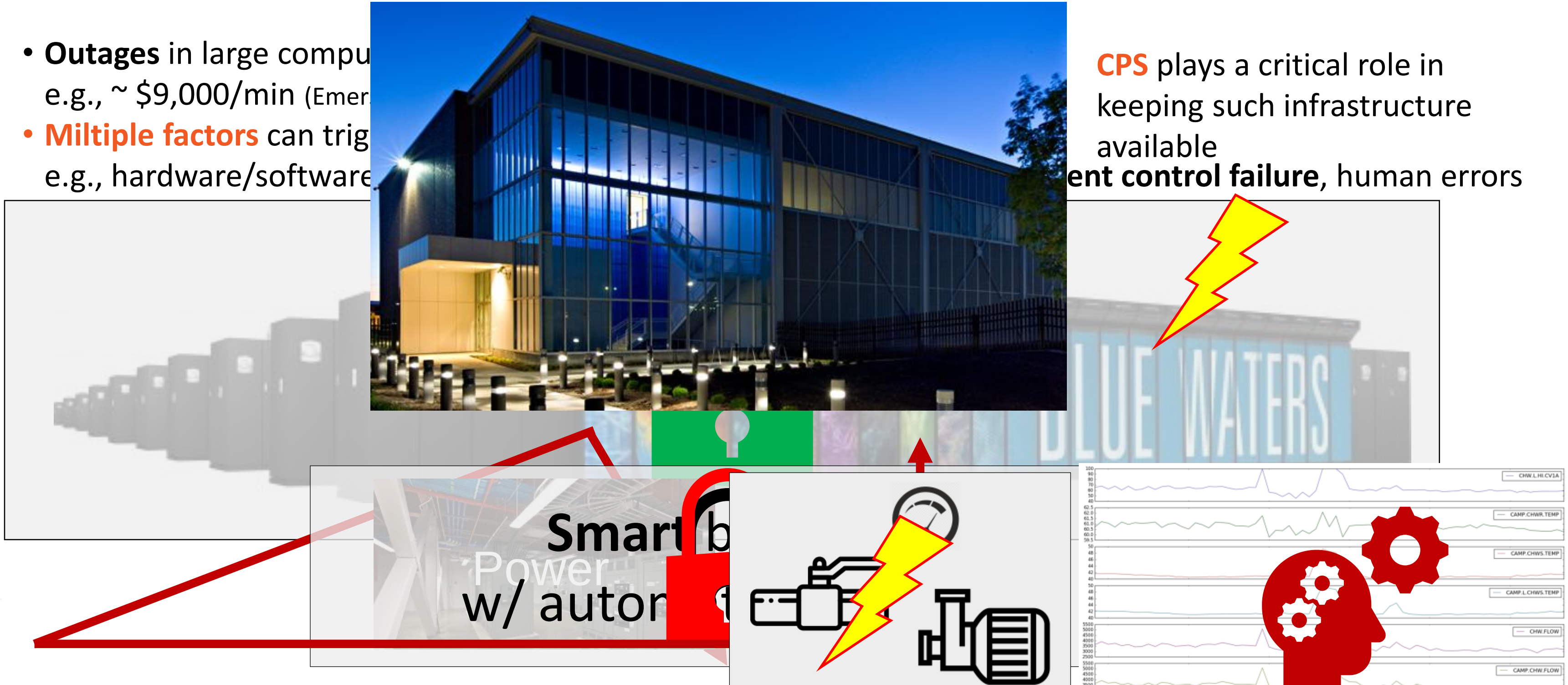
- *What Changed in Recent Years?*
  - Targeted Attacks evolved as a highly evasive attacks powered by artificial intelligence (AI) -> **AI-Driven Smart Malware**

[illegible] Large computing infrastructure are tempting targets for adversaries [illegible]

- **Outages** in large compu
  e.g., ~ $9,000/min (Emer
- **Miltiple factors** can trig
  e.g., hardware/software

CPS plays a critical role in keeping such infrastructure available

ent control failure, human errors



UPS failure in UK data center grounded 75K passengers. Loss of $106M (*2017*)

Cooling failure in MS data center resulted in 21hrs outage (*2018*)

Smart Malware to Bring Down Computing Enterprise

- **Indirect attack**
  - An attacker exploits ***relatively weak security*** of a CPS that manages the environment in which a major computing enterprise (e.g., HPC system or cloud infrastructure) operates
  - CPS is often ***outside the monitoring range*** of security monitoring deployed in the computing infrastructure
- **Stealthy attack**
  - Intruders ***masquerade an attack as an accidental failure*** in the CPS to mislead operators
  - Attackers may ***remain long time in the CPS*** without being noticed
- **Smart malware**
  - ***Dynamically infer*** (based on CPS operational data) attack strategies that mimic behavior corresponding to an accidental failure in the CPS

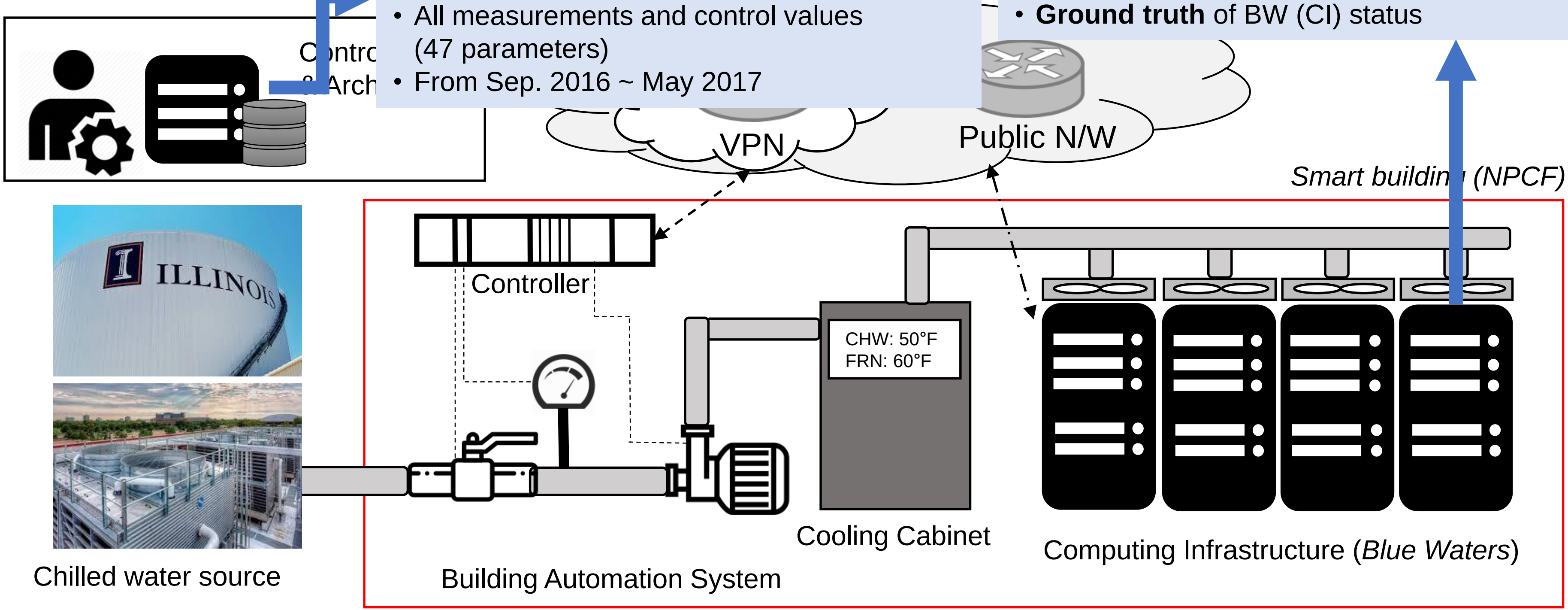
System & Data Overview

CPS operational data set:

- For ***analysis of CPS operation*** and ***inference of failure related information***
- All measurements and control values (47 parameters)
- From Sep. 2016 ~ May 2017

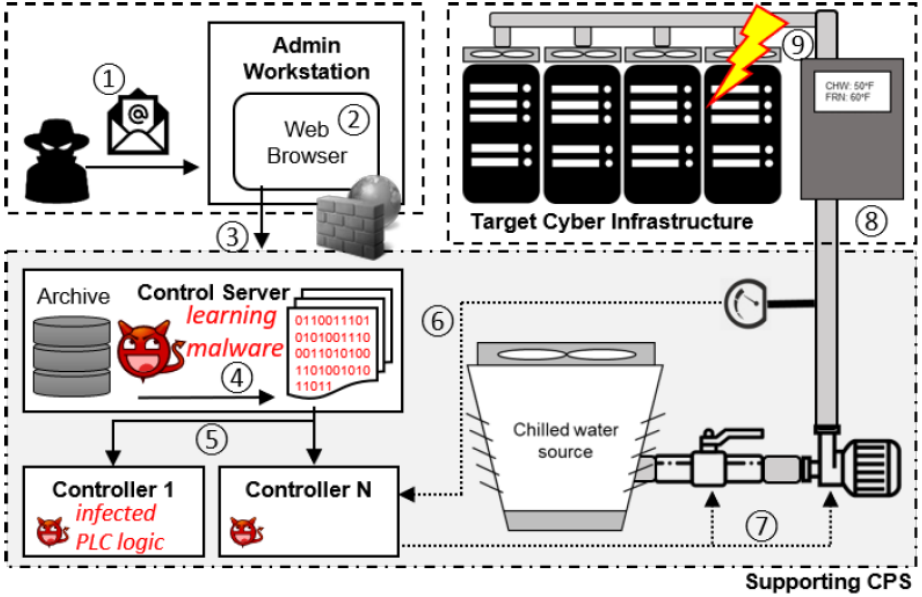
Blue Waters Incident reports:

- For **verification** of the prediction/inference (not available to attackers/malware)
- **Ground truth** of BW (CI) status



Smart Malware Approach

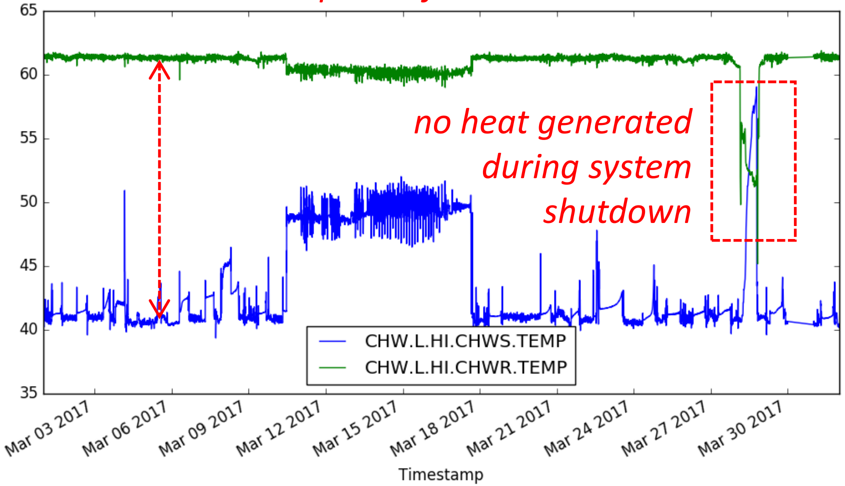
- *Steps 1 – 3:* initial compromise and establishment of a foothold in the target (up to installation)
- *Step 4:* cyclic sequence of procedures in reconnaissance and customization.
- *Step 5:* lateral movement into the physical control layer of the CPS
- *Step 6:* collection of information to evaluate the triggering condition
- *Steps 7 – 9:* acting on objective



Filtration of Failure Data **(Step B)**

- Attacker wants to ***masquerade*** an attack as an accidental failure
- Needs data on computing infrastructure (CI) failures and corresponding CPS events
- But! **No knowledge on CI status only CPS data**
- **Observation**:
  Chilled water return temperature constant
  Nadirs indicate *less heat absorbed from CI*

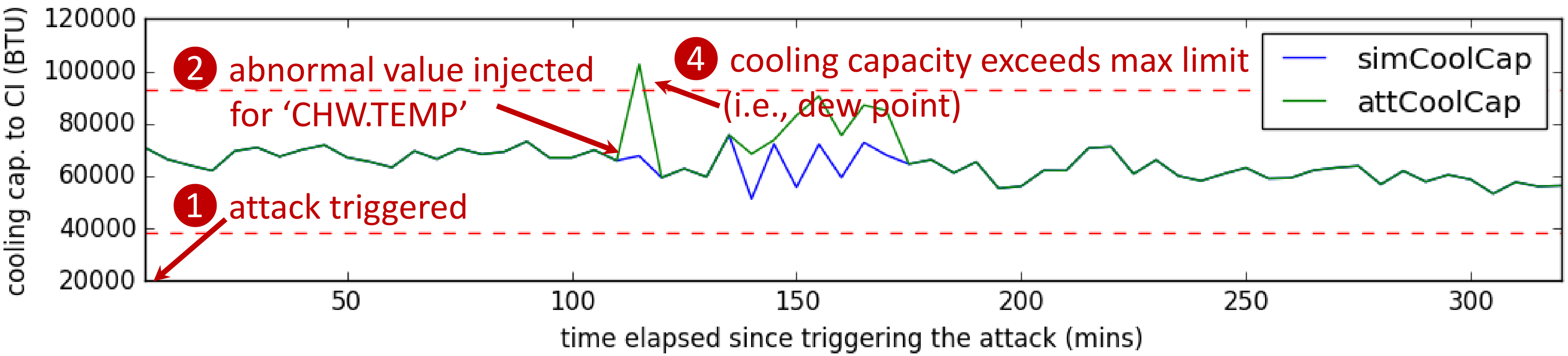
abstraction of the heat generated by the compute infrastructure



no heat generated during system shutdown

Attack Strategies Simulated

- **Smart malware inferred three CI outage-related strategies from CPS data**
  - ○ Supply water **temperature abnormality** due to power interruption
  - ○ **Chilled water loop closure** for building maintenance operation
  - ○ **Reduced cooling capacity** for emergency outage in computing infrastructure



1 attack triggered

2 abnormal value injected for ‘CHW.TEMP’

4 cooling capacity exceeds max limit (i.e., dew point)

3 In response to the fake increase in temperature, CPS-simulator increases flow (i.e., opens the valves)

Conclusions

- Self-learning Smart Malware no longer a remote possibility
  - Its success depends on the availability of the data

- Presented example of **smart malware** (smart building)
  - Generalizable to other systems that employ similar design & architecture
  - Strategies not hard-coded into malware – dynamically derived from data

- Protection against smart malware:
  - ***Supervised-learning driven detectors***
    - take advantage of knowledge on the runtime status of the control infrastructure and details of the CPS available to the operators
  - ***Multi-layered monitoring***: deploy monitors in the physical layer in addition to the cyber layer (e.g., dedicated IDS)

Acknowledgments

- *Collaborators:*
  Ph.D. student: K. Chung

  Prof. R. Iyer

- *Project sponsors:*
  - National Science Foundation (NSF)
  - National Center for supercomputing Applications (NCSA)