Digital Twin-Based Cyber-Attack Detection Framework for Cyber-Physical Manufacturing Systems

서울과학기술대학교 컴퓨터공학과 진호천

Depart. Computer Science, Seoul National University of Science and Technology

CONTENTS

IV.C Integration of the DTs for Attack Detection

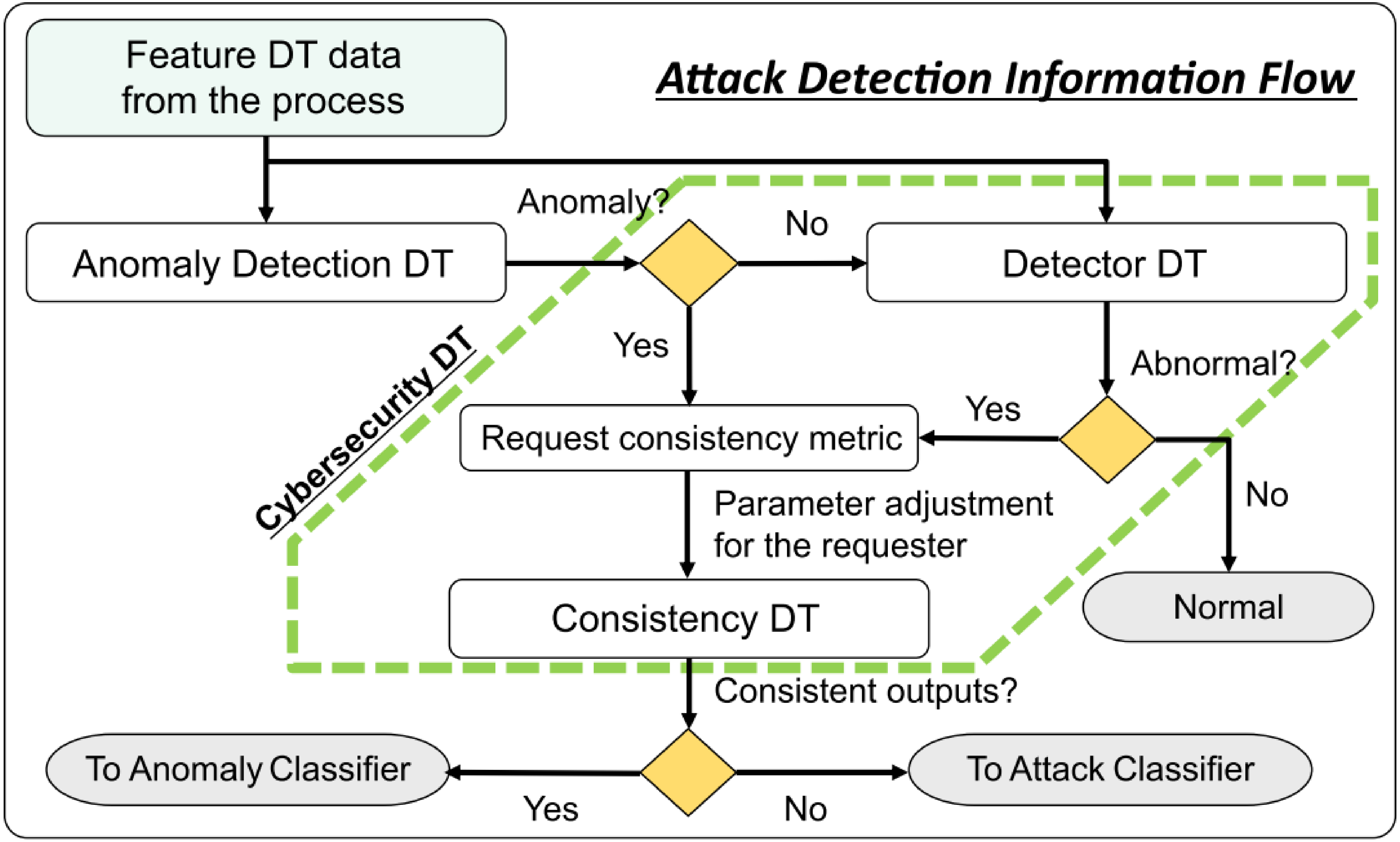
V. EXPERIMENTAL DEMONSTRATION ON AN OFF-THE-SHELF 3D PRINTER

- Controller DT Over a Network
- Attack and Anomaly Scenarios
- Cybersecurity DT
- Implementation Details
- Results

VI. Conclusion

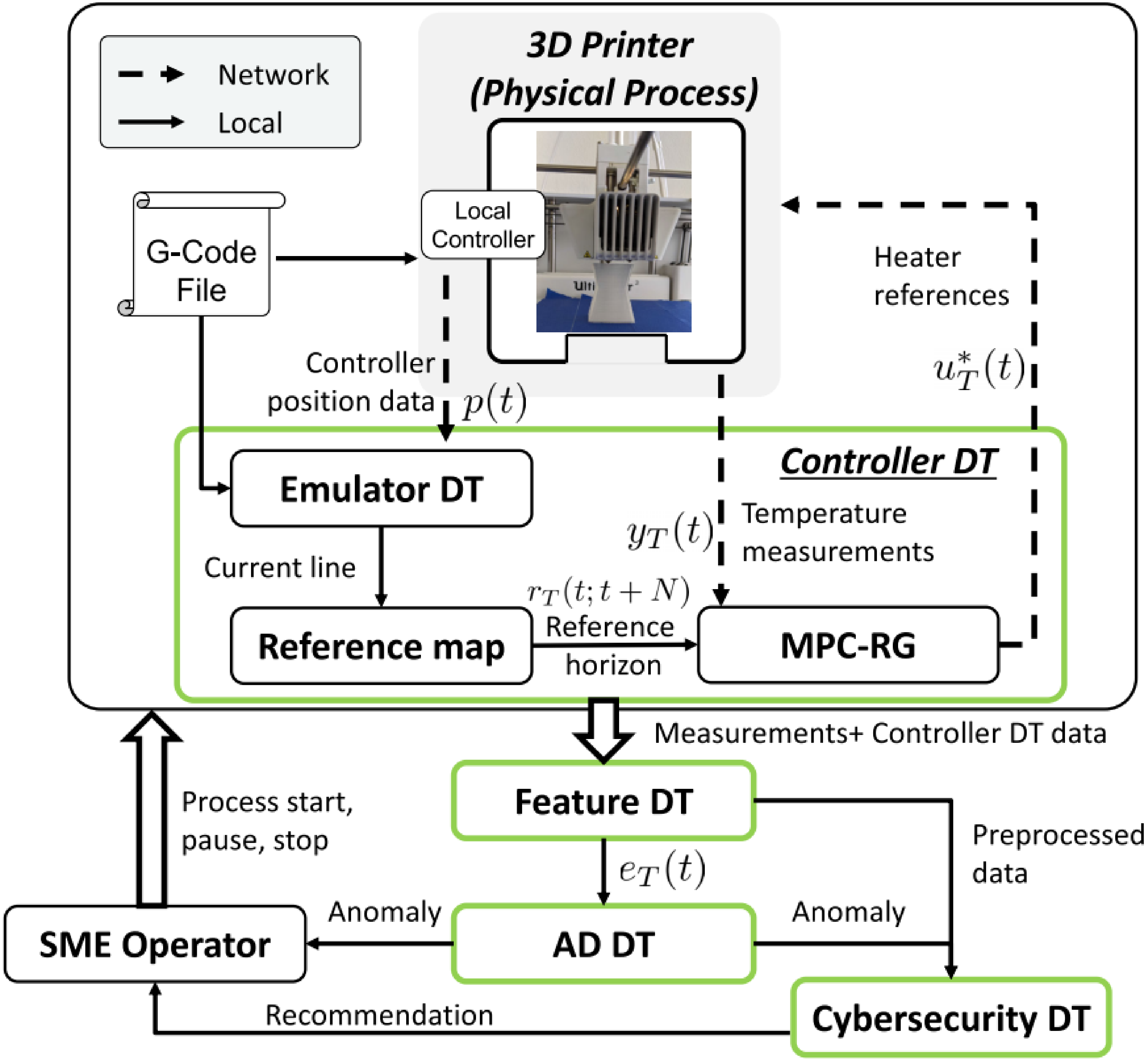
Etc. My Opinion

Integration of the DTs for Attack Detection



Experiment Environment

By means of network security DT, the experimental data of normal operation, abnormal operation and attack of the printer are collected and analyzed, and the attack detection results are obtained. The experimental setup is summarized in the figure below. The experimental study is a demonstration of the framework presented in a real world setting. The framework can be used in a variety of application scenarios, not just those described here, and can complement existing methods.



Controller DT Over a Network

- **Motivation**: Heating the deposited material within the desired temperature range is crucial for an extrusion process.
- **Controller Implementation**: Since we do not have direct access to the nozzle heaters in the printer, we model the closed-loop heating system and develop a model predictive controller (MPC) scheme to prescribe heater references so that the system output y_T (t) tracks a reference temperature r_T (t).
- **Reference Handling**: G-Code references executed on the printer are inherently spatial and event-based. To remedy this mismatch, authours utilize an Emulator DT that emulates the printing process by analyzing the G-Code file.

$$\boldsymbol{x}(t+1) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}u_T(t) \tag{7a}$$

$$y_T(t) = \boldsymbol{C}\boldsymbol{x}(t), \tag{7b}$$

$$\min_{\boldsymbol{u}} \sum_{\tau=t}^{t+N-1} ||\boldsymbol{x}(\tau) - \boldsymbol{x}^r(\tau)||_Q^2 + ||u_T(\tau) - u_T^r(\tau)||_R^2 \tag{8a}$$

$$+||\boldsymbol{x}(t+N) - \boldsymbol{x}^r(t+N)||_P^2 \tag{8b}$$

$$\text{s.t.:}\quad \boldsymbol{x}(\tau+1) = \boldsymbol{A}\boldsymbol{x}(\tau) + \boldsymbol{B}u_T(\tau) \tag{8c}$$

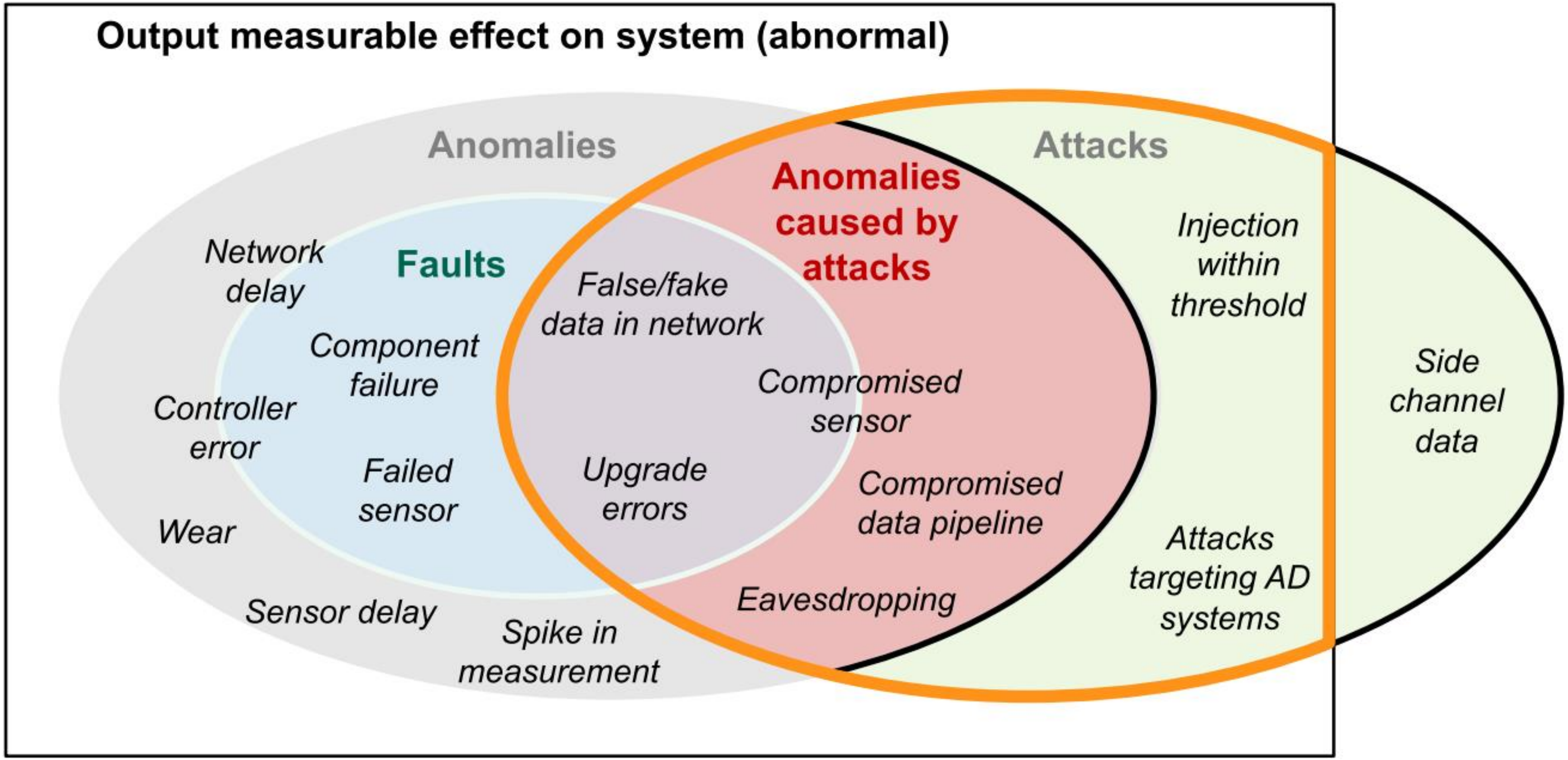
$$\boldsymbol{x}(t) = \hat{\boldsymbol{x}}(t),\ u_T(\tau) \in \mathcal{U},\ \tau = t, \ldots, t+N-1 \tag{8d}$$

Attack and Anomaly Scenarios

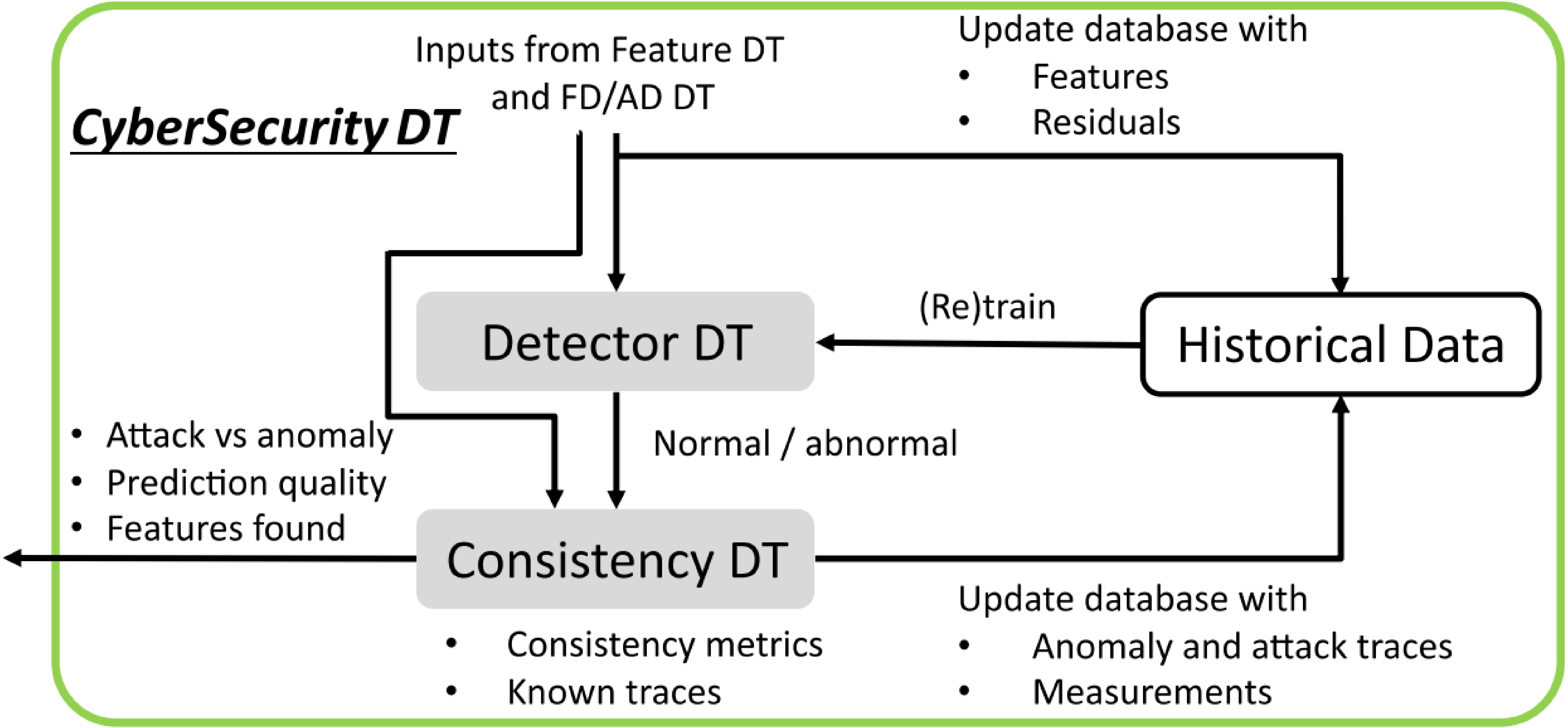
- **Anomalies**
  - The first anomaly in the extruder head temperature measurement is caused by the presence of a cooling fan. The fan increases the extruder nozzle's airflow, which lowers its temperature, causing an unknown disturbance that is not accounted for by the controller.
  - The second anomaly is due to the gradual degradation of the heating system's performance. Over time, the system's components experience wear and tear, resulting in a slower response time than expected for a given temperature reference r_T (t). This effect occurs gradually over several months of use and is simulated by updating the local controller gains to slow down the closed-loop response of the heating system intentionally.
- **Attacks**
  - Injection of a constant offset to the measurement signal $w(t) = c_1$ for some $c_1 \in \mathbb{R}$.
  - Injection of a temporally cyclic signal to the measurement signal $w(t) = c_2 \sin(t)$ for some $c_2 \in \mathbb{R}$.

Attack and Anomaly Scenarios

- **Remark**: Authors do not differentiate attacks based on their malicious or non-malicious intent.



Cyber Security DT



CyberSecurity DT

Inputs from Feature DT and FD/AD DT

Update database with
- Features
- Residuals

- Attack vs anomaly
- Prediction quality
- Features found

- Consistency metrics
- Known traces

Update database with
- Anomaly and attack traces
- Measurements

Detector DT:

$$\boldsymbol{\alpha}^* = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\{\boldsymbol{\alpha}^T \boldsymbol{K}\boldsymbol{\alpha} \mid 0 \le \alpha_i \le \frac{1}{\upsilon n_z}, \textstyle\sum_i \alpha_i = 1\}, \tag{9}$$

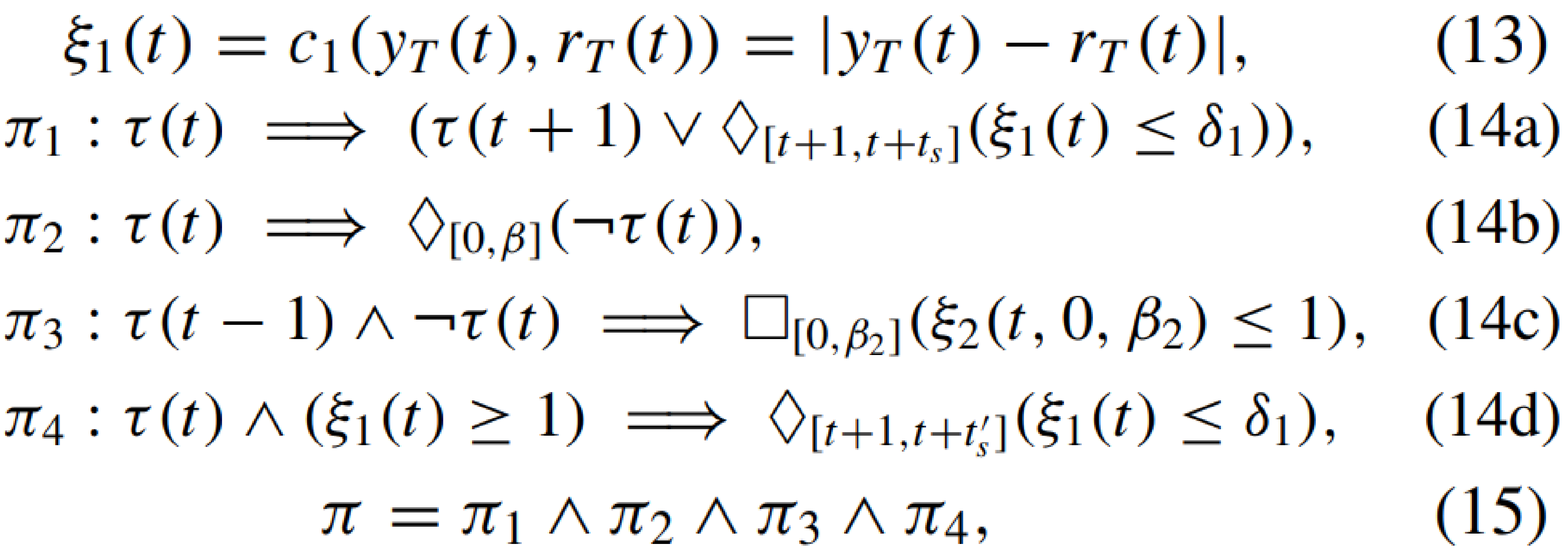
$$\rho^* = \textstyle\sum_i \alpha_i k(z_j, z_i). \tag{10}$$

$$h(z_*) = sgn(\textstyle\sum_i \alpha_i k(z_i, z_*) - \rho^*). \tag{11}$$

$$\mathcal{B}(D^+) = \{z \mid h(z) \ge 0\}. \tag{12}$$

Remark: The abnormality detection checks the condition $y_T\ (t - n_{sp}) \in B_i$. If the volume of B_i is too large (in a multidimensional sense), projections of certain attacked process measurements may still be within B_i , resulting in false negatives.

Consistency DT:



$$\xi_1(t) = c_1(y_T(t), r_T(t)) = |y_T(t) - r_T(t)|, \tag{13}$$

$$\pi_1 : \tau(t) \implies (\tau(t+1) \vee \Diamond_{[t+1,t+t_s]}(\xi_1(t) \le \delta_1)), \tag{14a}$$

$$\pi_2 : \tau(t) \implies \Diamond_{[0,\beta]}(\neg\tau(t)), \tag{14b}$$

$$\pi_3 : \tau(t-1) \wedge \neg\tau(t) \implies \Box_{[0,\beta_2]}(\xi_2(t, 0, \beta_2) \le 1), \tag{14c}$$

$$\pi_4 : \tau(t) \wedge (\xi_1(t) \ge 1) \implies \Diamond_{[t+1,t+t'_s]}(\xi_1(t) \le \delta_1), \tag{14d}$$

$$\pi = \pi_1 \wedge \pi_2 \wedge \pi_3 \wedge \pi_4, \tag{15}$$

Implementation Details

Equipment: Ultimaker 3 FFF printer with Network-API

Monitoring of extruder temperatures: y_T (t)

Stepper motor counts: p(t)

Extruder heater input: u_T (t)

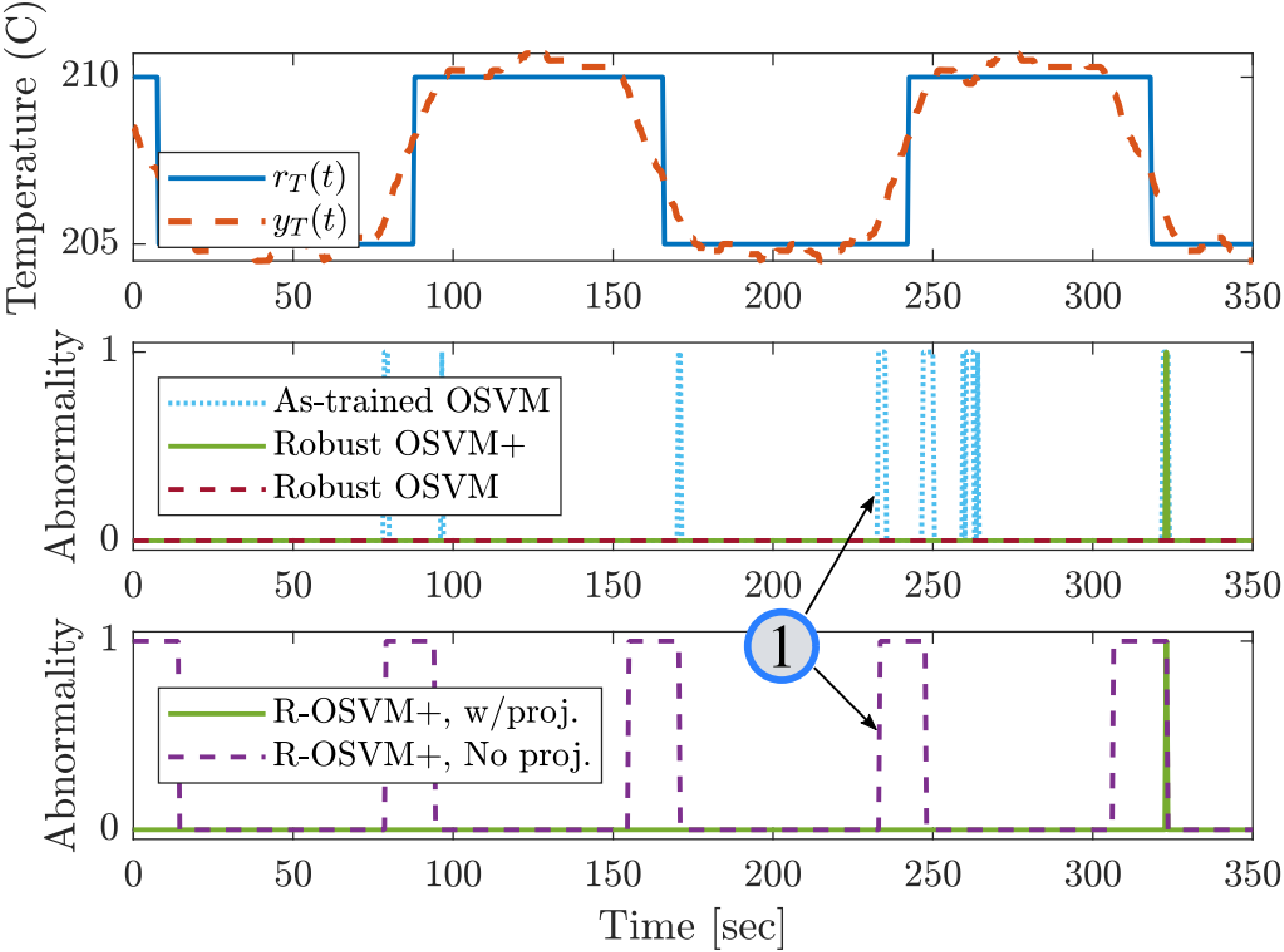
New heater reference: $u_T^* (t)$

Time-series data:

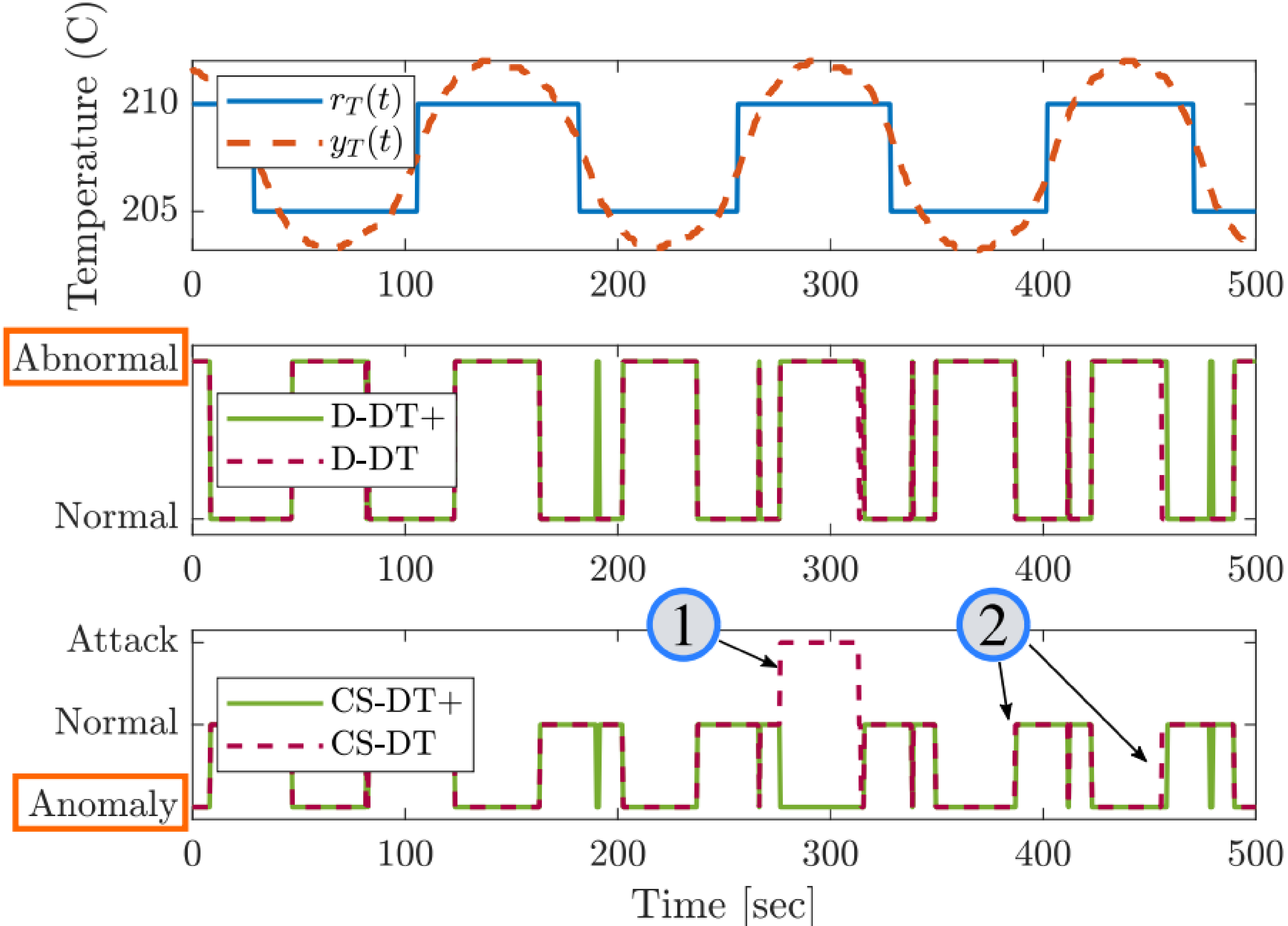
- the nominal system without the Controller DT running
- the controlled system with the Controller DT without attacks or anomalies
- the controlled system with anomaly cases (A1and A2), and (iv) the controlled system with sensor attack cases (T1, T2)

Analytical Method: Authors conduct data analysis in an asynchronous offline manner. This approach is taken to avoid any computational errors in our results due to possible runtime execution issues, and to illustrate how various DTs in the framework may operate at different time scales. It is important to note that synchronous implementations of the Cybersecurity DT would perform comparably in practice to our approach since we do not change how the Cybersecurity DT interacts with the other DTs in the framework in our experimental approach. The experimental data used in this case study is available in the supplementary materials.

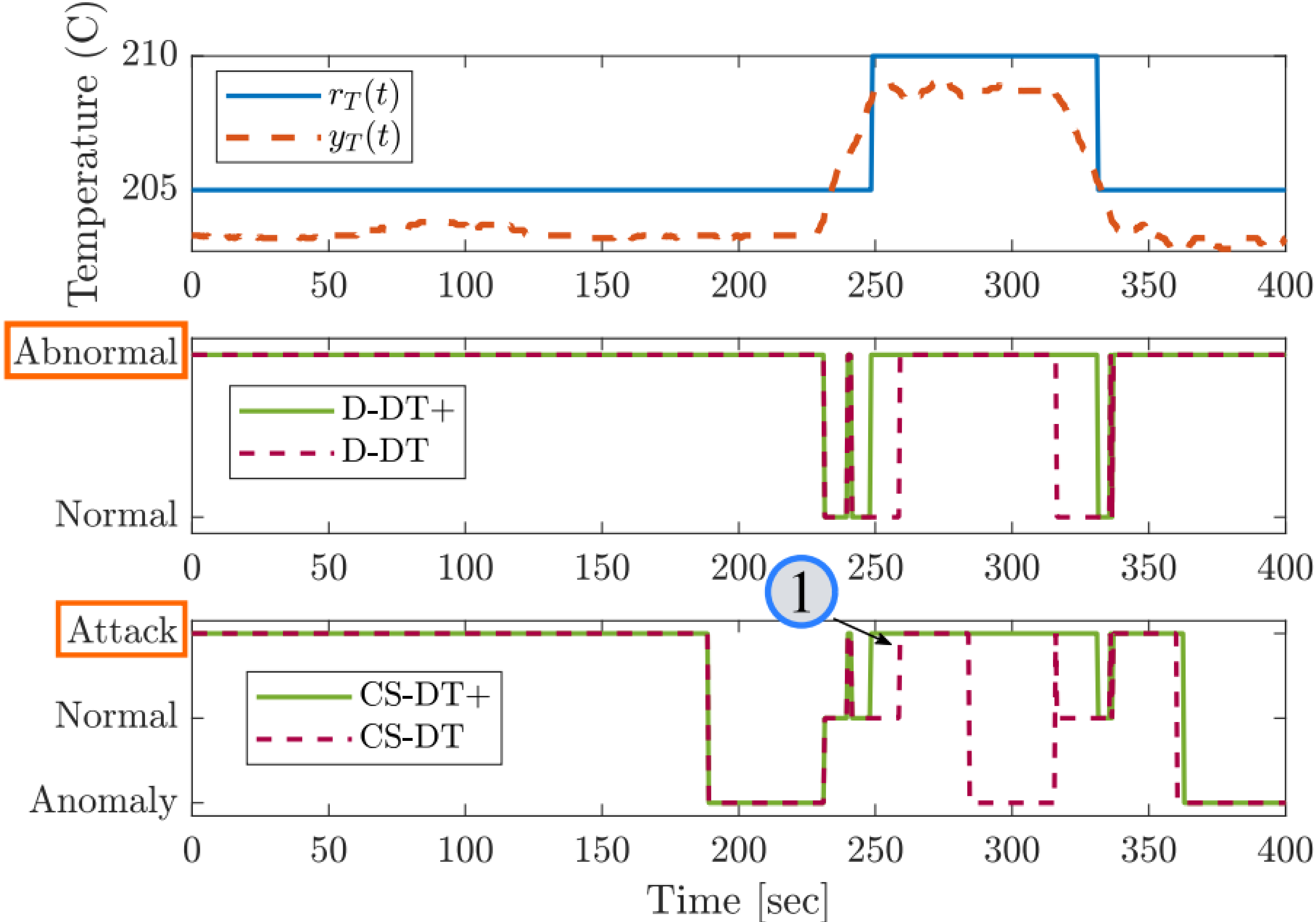
Results-Performance of Detector DT



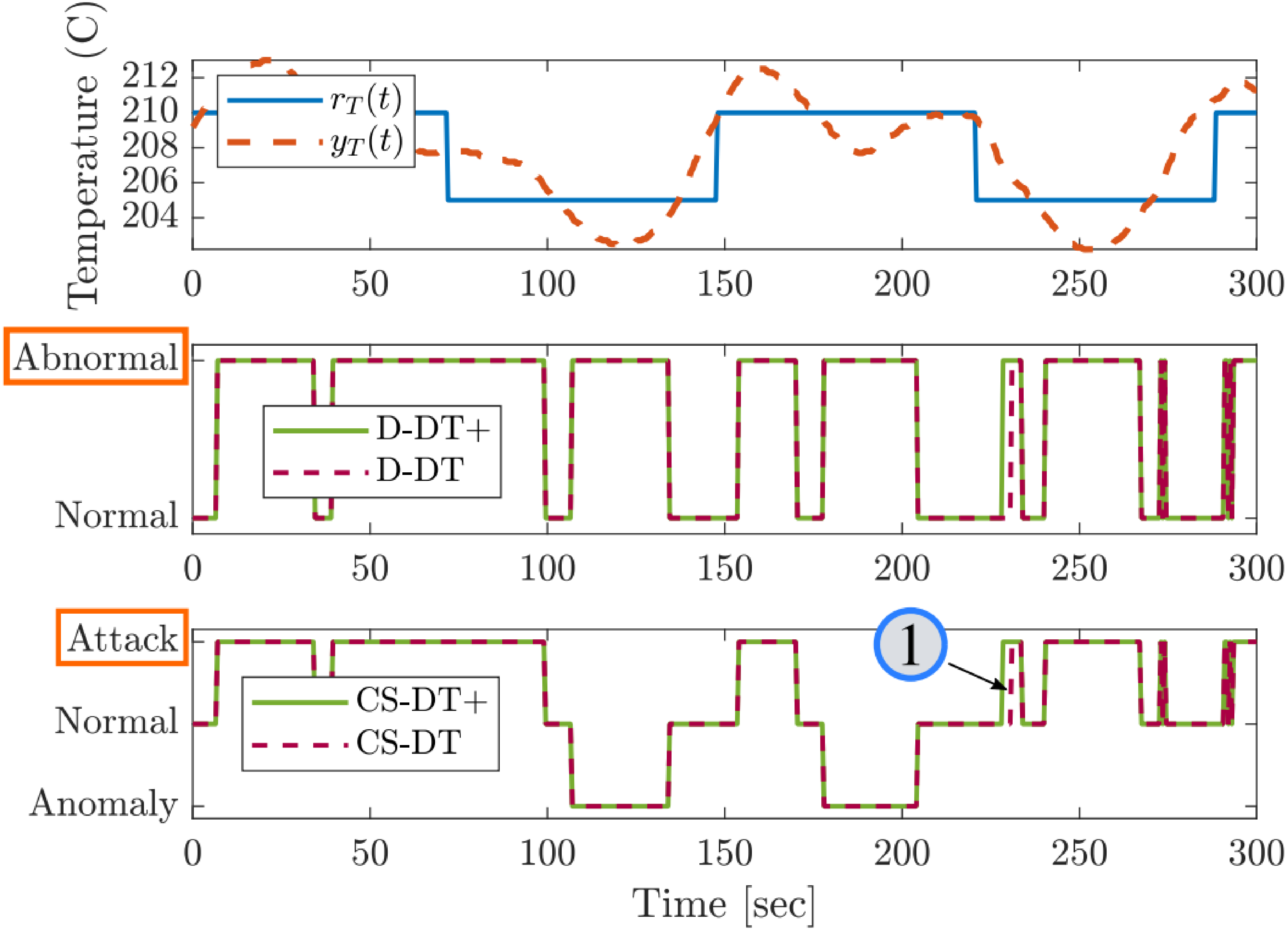
Results-Performance of Security DT



Results-Performance of Security DT



Results-Performance of Security DT



Discussion

The results of this section show that the Cybersecurity DT can identify inconsistencies in the signal and predict potential attacks. It can also correctly analyze and identify attacks without adjusting parameters or retraining the model.

However, there are still some missed or falsely identified signals, especially around setpoint changes, due to the trade-offs in the framework design. The goal is to provide robustness during normal processes and reduce false alarms. The design aims to minimize false alarms during the printing process, where the cost of stopping prematurely can be high. In this case study, the Cybersecurity DT uses an OSVM detector with STL consistency checks. The modular and flexible design allows for the integration of other methods as needed. The framework demonstrates flexibility and generality for developing DTs tailored to specific needs. The observation that incorporating latch-up detection can improve detection performance is important. The framework can be extended to develop further methods for analyzing transitions between states, which may require additional modeling and data analysis.

Outlook for Additional Application Scenarios

Designing detection and consistency DTs requires knowledge in physics and SME, as well as suitable sensors. However, in many practical applications, such as a wide range of physics-based detection methods, this is not a limitation. Additionally, consistency features can be learned from past data using data-driven classification and pattern matching methods, reducing the need for SME knowledge in designing consistency DTs. Assuming necessary sensors are available in the given CPMS, the study focuses on measurable attacks outputs.

The paper [24] studied product-oriented attacks in machining processes and used in-situ monitoring of process variables for attack detection. In authors framework, measurement systems can be implemented in the physical process or via the Controller DT, with data processed by the Feature DT to evaluate relevant metrics and shared with the Cybersecurity DT. The Cybersecurity DT may implement attack detection logic presented in [24], with the Consistency DT using STL logic to monitor parameter changes within SME-specified limits.

Outlook for Additional Application Scenarios

In [26], data-driven methods for detecting cyber-physical attacks are proposed and demonstrated on AM and machining processes. The Feature DT can process the measurement data streams to prepare features for analysis by the Cybersecurity DT. The data-driven methods proposed in [26] can be implemented as part of the Detector DT, similar to the OSVM application shown in our work. From the case studies in [26], images are collected on the 3D printer by a microcontroller and streamed to the framework via the Feature DT. The images are then logged to a database, and feature signals such as mean, standard deviation, and magnitude of pixel values are evaluated and shared with the Cybersecurity DT. Data-driven models are utilized to detect cyber-attacks and alert an SME. However, methods to differentiate anomalies from attacks and dealing with advanced closed-loop controllers have not been discussed in [24] and [26].

Other works from the literature that utilize side-channel measurements or digital signature matching methods can be implemented within our framework using appropriate Detector and Consistency DTs. The data processing and attack detection methods in [29] can also be implemented with Feature and Detection DTs. Similarly, SPC methods and control charts can be integrated into the Detector DT to find statistically significant variations, which may be further analyzed by the Consistency DT. Our framework further enables tools to deal with a large class of attack types and the use of advanced control methods with transient behavior. However, a detailed study of implementing the full set of SPC tools within our framework is beyond the scope of this work and subject for future research.

Conclusion

The article describes a framework for cyberattack detection on cyber-physical manufacturing systems (CPMS) in the context of closed-loop controllers and anomalies in the physical process. The proposed Cybersecurity DT can detect attacks and anomalies on the system while the process is controlled to switching setpoints. The framework is platform agnostic and modular, and its components can be extended for multiple resources in a manufacturing system and aggregated into a system-level DT framework. Future work includes investigating cases where anomalies and attacks are present in the system simultaneously and extending the presented abnormality detection approach as a stand-alone method for CPMS. Additionally, studying attacks with no output measurable effects is of interest for future work.

Opinion

This paper presents a comprehensive framework and system, including theory, practical deployable experiments, and combinations with other possibilities. It is a highly complete paper.

The drawback of this paper is the lack of differentiation between malicious and non-malicious attacks based solely on information capture, leading to a significant false positive rate.

Thanks

서울과학기술대학교 컴퓨터공학과 진호천

Depart. Computer Science, Seoul National University of Science and Technology