Provable Adversarial Safety in Cyber-Physical Systems

John H. Castellanos
CISPA Helmholtz Center for Information Security
Saarbrücken, Germany
john.castellanos@cispa.de

Mohamed Maghenem
University of Grenoble Alpes, CNRS
Grenoble INP, France
mohamed.maghenem@gipsa-lab.grenoble-inp.fr

Alvaro A. Cardenas
University of California
Santa Cruz, USA
alacarde@ucsc.edu

Ricardo G. Sanfelice
University of California
Santa Cruz, USA
ricardo@ucsc.edu

Jianying Zhou
Singapore University of Technology and Design
Singapore
jianying_zhou@sutd.edu.sg

***Abstract*—Most proposals for securing control systems are heuristic in nature, and while they increase the protection of their target, the security guarantees they provide are unclear. This paper proposes a new way of modeling the security guarantees of a Cyber-Physical System (CPS) against arbitrary false command attacks. As our main case study, we use the most popular testbed for control systems security. We first propose a detailed formal model of this testbed and then show how the original configuration is vulnerable to a single-actuator attack. We then propose modifications to the control system and prove that our modified system is secure against arbitrary, single-actuator attacks.**

1. Introduction

In the past decades, we have seen several confirmed attacks on industrial control systems, including a sewage control system in Australia [56], a nuclear enrichment facility in Iran [68], the power grid of Ukraine [13], a steel mill in Germany [39], a paper mill in Louisiana [6], oil systems in the Middle-East [35], and a water utility in Florida [44]. In all these cases, an attacker partially compromised a control system and then sent malicious control commands to the physical process, causing accidents and damages.

Researchers have suggested various defense strategies; however, two main challenges remain largely unaddressed. First, most security efforts for industrial control security are heuristic in nature, and they do not provide provable security assertions about the system's safety. Second, implementing and evaluating security proposals for industrial control systems is generally restricted to either simulations or toy physical systems, given the difficulty of getting access to real-world operational industrial plants.

This paper addresses these two limitations by proposing a new method to formally prove safety properties against an attacker that has partially compromised the system and implementing and testing our formal model in a real-world operating plant. In particular, we implement our proposal in the most popular industrial process for CPS security research [15]: the Secure Water Treatment testbed (SWaT).

First, this paper proposes the most comprehensive formal mathematical model of SWaT. As we show in Table 1, there is no previous model that is larger than ours (our representation of SWaT captures in a single model as many or more SWaT elements than previous work). We have also released our model as open-source software (see itemized contributions below). In addition, while several papers in top security conferences have used SWaT before [3], [9], [12], [21], [62], none of these previous efforts attempted to provide a comprehensive formal model of the system nor enable proofs of security assertions.

We then formally prove the following sequence of results: (1) We prove that without attacks, SWaT is safe. (2) We show that if the attacker can compromise a single actuator in the system, then the original design of SWaT is unsafe (some attacks drive the system to unsafe regions). (3) We propose modifications to the Programmable Logic Controllers (PLCs) operating SWaT and then formally prove that SWaT will remain safe under single actuator attacks with these changes. (4) We show that if the attacker can compromise two or more actuators, SWaT is unsafe, regardless of the logic in the PLCs.

To prove these security assertions we extend the theory of barrier function certificates for hybrid systems and adapt them to analyze the safety of a system under attack. Our contributions include extensions to the theory of barrier functions by introducing the concept of uniform safety in the presence of arbitrary exogenous signals.

Furthermore, we propose a new adversary model that does not make parametric assumptions about the attacker's tactics. Our adversary model only needs to know the number of actuators an attacker has under its control but does not need to know the tactics. In other words, the attacker can launch a square wave attack of any frequency, a delay attack, an inversion attack, etc. Any arbitrary attack signal is considered in our model, making our security proofs robust to new unanticipated attack tactics. This is contrary to most work on industrial security which makes parametric assumptions about the attack strategies (e.g., scaling attacks [59], bias attacks [12], [14], delay attacks [59], or random attacks [20], [66]). We note that actuators can be compromised individually by targeting the Remote Input/Output (RIO) computer interfacing the actuator with the PLC [63].

Finally, we evaluate and test our methods and proofs in a real-world system. We also show evidence that our model is accurate by comparing our model and traces from the real-world system.

In summary, our contributions include:

1) We introduce the most comprehensive formal model of the most popular [15] testbed for CPS security.
2) We formally prove several security assertions about SWaT, including (1) safety without attacks, (2) unsafe if an attacker compromises a single actuator, (3) safe with our PLC modifications (under a single actuator attack), (4) unsafe with two or more actuators under attack.
3) We extend the theory of barrier function certificates for hybrid systems by introducing a new concept for proving safety in adversarial conditions. We then find sufficient conditions to prove safety under our adversary model.
4) Our adversary model considers attackers that can launch arbitrary control commands. Our adversary model is more realistic and powerful than previous adversaries considered in industrial control systems [12], [14], [20], [21], [59], [66].
5) We implement and validate our approach in a real-world system. We release our model as open-source software[1] in the hopes that other researchers working with SWaT can use it.

1.1. Organization of the paper

The rest of the paper is organized as follows: In Section 2 we discuss related work. Then in Section 4 we introduce the mathematical tools to model a cyber-physical system, our adversary model (Section 4.1), and a theorem to show how to prove the safety of control systems under attacks. In Section 5 we introduce the physical model of SWaT. We also prove that this system is safe without attacks for the first time. We evaluate the security of our water system in Section 6 and show how the system is safe under some attacks and unsafe under other attacks. We then propose changes to the control logic and to the physical parameters of the system to make the system safe, irrespective of any individual control signal being compromised. We then finalize our analysis with experimental results validating our theoretical model in Section 7.

2. Related Work

2.1. Attacks to CPS

There are various ways in which attackers can take over sensors or actuators in control systems. In addition to compromising devices in the "classical" way (e.g., a software exploit), attackers can also compromise sensor or control signals with novel physical attacks. Analog sensor security [67], [23], [61], [57], [8] focuses on how physical interference can affect the reported sensor readings back to the control system. By adding a physical signal (electromagnetic, sound, heat, etc.), attackers can affect sensor readings. In some cases, these types of attacks can also manipulate actuators directly [54], [16].

When attackers compromise a control signal in a CPS (e.g., acceleration in a vehicle), they inject a time series to the physical system $a(t)$, where t denotes time. Researchers tend to parameterize the adversary tactics to fixed strategies, and they assume that the attacker simply replaces the non-compromised signal $u(t)$ with a parameterized version of it. Examples include scaling attacks $a(t) = \alpha u(t)$ [59], bias attacks $a(t) = u(t) + b$ [12], [14], delay attacks $a(t) = u(t - d)$ [59], and random attacks (where $a(t)$ is a random value at each time) [66], [20].

As history has taught us, limiting the attacker to follow specific attacks will not guarantee security. Over the years, system after system has been defeated by adversaries that break the assumptions of the model. A recent high-profile example is the case of attacks against key handshakes in WPA2 [64] which were proven secure [30] under a model that did not capture key installation. To increase the confidence of a security proof, adversary models used in formal proofs tend to be as general as possible [17], [37], [27], e.g., by assuming adversaries to be **any** polynomial time algorithm [27], [37] (without parameterizing the specific attacker algorithm used to crack the system). Our goal in this paper is to study attacks without assuming a priori a fixed attack tactic.

2.2. Securing CPS

Defenses for control systems can be reactive or proactive. Reactive security focuses on detecting attacks [12], [14], [21], [62], [29], [53], [50] and sometimes responding to attacks [69], [50], [20], [19]. Attack response usually focuses on identifying the malicious sensor or actuator signals and eliminating them.

Proactive security proposals, on the other hand, focus on secure design: for example, designing a control algorithm so that the system is more resilient to attacks or designing actuators so what the attacker can do is limited [36], [24]. This paper focuses on proactive security: *we want to evaluate the system offline to understand the impact of attacks and if possible, redesign the controller to minimize the negative impacts of any future attack.*

In summary, the **scope** of this paper is to study (offline) the safety of an industrial design, identify the stress points where an attacker can break the system., and propose improvements to make the system more secure and resilient to attacks. We approach this problem by providing provable security assertions.

2.3. Formal Verification of CPS

Embedded control systems monitor and control a variety of safety-critical problems. To formally guarantee the safe operation of CPS, we need formal models and rigorous verification approaches. In this subsection we summarize different ways to formally verify a CPS and introduce how the tool we use in the paper (barrier certificates) compares to alternatives.

In most classical problems in computer science, model checking focuses on discrete dynamics; however, the

[1] https://gitlab.com/cpsResearcher/watertmodel

unique challenge of verifying CPS properties arises due to their continuous dynamics. Continuous dynamics creates new problems as we need to consider the evolution of physical states that follow trajectories defined by differential equations.

There are three main approaches for the verification of CPS: (1) set-based reachability analysis, (2) abstraction-based verification, and (3) logic-based verification [18].

Reachability analysis attempts to find the set of reachable states of a CPS as it evolves over time. The goal is to check if the reachable set and the unsafe set are disjoint. There are several tools for reachability calculations; for example, Flow* [11] uses a flowpipe construction scheme to verify time-bounded reachability. Reachability analysis has also been applied in an adversarial setting [69], [36], [24], [58]; however, these security efforts for applying reachability in adversarial conditions have only considered linear systems, while real-world CPS have more complex physical behaviors (like hybrid dynamics). In this paper, we go beyond linear systems and apply our methods to a real-world control system modeled by hybrid equations.

Abstraction-based verification attempts to address the scalability problems of reachability analysis. To scale up discrete model checking, abstraction-based verification replaces the actual system with a simpler, abstract system in which model checking is easier to perform [31], [4]. The drawback of these abstractions is that any verification result can only be related back to the original system if the property in question survives the abstraction process [18].

Finally, logic-based verification provides a witness to verify a continuous system respects the desired property. There are two main methods for logic-based verification: (a) differential invariants [45], [46], and (b) barrier certificates [47], [5]. Differential invariants are based on Lie derivatives and Lie groups, while barrier certificates are based on Lyapunov's criterion for stability [18]. Differential invariants are the most general representation for logic-based verification, but if you find a barrier certificate, logic-based verification can be proven directly (analytically) and without computational support (e.g., requiring software tools). In addition, software tools require the discretization of continuous states, while analytical barrier certificate proofs do not require this approximation [33].

Barrier certificates are a way to separate good and bad states and to show that this separation (barrier) is impenetrable by the continuous system dynamics. The importance of barrier certificates comes from the fact that they reduce a reachability question (can we ever reach an unsafe state) by a simple check on the directional derivative of the barrier certificate along the differential equation of the system [18]. In contrast with other tools, barrier certificates provide formal analytical guarantees without extra assumptions or without the need to rely on computational tools. For example, reachability tools like Flow* [11] use a flowpipe construction scheme to verify time-bounded reachability. Flowpipe construction methods are often easily used since users only need to specify the flow pipe stepsize, approximation order, and the bounded time horizon. On the other hand, Barrier certificates require an expert to find a barrier function, but a barrier certificate can be used to prove time-unbounded reachability. In short, Barrier certificates, like the ones we obtain in this work, make safety proofs harder, but their safety guarantees are stronger than time-bounded reachability alternatives, proofs relying on abstractions, or proofs relying on bounded computational checks.

In this paper, we develop the theory of barrier certificates for CPS verification of safety under actuation attacks. We also contribute to the literature on safety by defining the new concept of *uniform safety*, which is required when an attacker is the source of uncertainty in the system. In addition, we prove new theorems that show how to check if a barrier certificate satisfies uniform safety.

In this paper we focus on formal verification, rather than testing. There have been other efforts to use testing for SWaT. For example, HyChecker [38] combines random sampling with symbolic execution on hybrid systems to perform probabilistic security testing. HyChecker is not a formal verification approach, so it cannot guarantee safety properties in the system. Barrier certificates can provide these safety guarantees.

2.4. Previous Models of SWaT

Previous works	**Modeling approach**	**AL**	**SWaT Stages Modeled**	**U**
Adepu [1], Feng [21]	Black-Box	C	Tank level, chemical flows	D
Ahmed [3], [2]	Black-Box	O	Sensor signals	D
Castellanos [10]	Black-Box	F	Tank level (Stage 1)	D
Chen [12]	Black-Box	F	Tank level (Stage 1, 3, 4)	D
Lin [40]	Black-Box	F	Tank level (Stage 1, 3, 4)	D
Urbina [62]	Black-Box	F	Chemical dosing (Stage 2)	D
HyChecker [38]	Probabilistic hybrid model	C	Backwash tank	T
This work	First-Principles	F	Tank level (Stage 1, 3, 4)	S

TABLE 1: SWaT models studied in previous security conferences. Abstraction level(AL): (O) Orthogonal model, (C) Coarse model, (F) Fine-grained model. Use case(U): (D) Attack Detection, (T) Security Testing, (S) Formal Proofs of Adversarial Safety.

SWaT is a real-world water system that has been widely used in security and formal method conferences [1], [2], [3], [9], [10], [12], [21], [38], [40], [62].

Some of these efforts have attempted to model SWaT. Table 1 shows how our model of SWaT compares to other models of SWaT. Our paper models the interactions of three interconnected stages in a single fine-grained mathematical model and therefore is the model capturing most of the physics in SWaT (other papers simply model a smaller subset of SWaT). Only two other related works [12], [40] have provided a similar comprehensive model of SWaT. Compared to them, our model is explicitly given by mathematical equations derived from first principles, while previous work attempting to capture the same level of complexity has resorted to black-box machine learning models that do not provide any guarantees about the accuracy of the model, or the explanation of the dynamics (therefore these previous models cannot be used for mathematical proofs of security).

Coarse models [22], [1] do not provide accuracy for the time series in the testbed and, therefore cannot model the precise effects of adversaries. And others [2], [3] are orthogonal models that do not attempt to model SWaT but instead attempt to fingerprint the innate noise of sensors.

In short, our work is the most complete model of SWaT available in the literature, and furthermore, it is a formal mathematical model that can be used in formal proofs of security, and it is based on first principles, so it can be used to explain the interactions between various components.

3. Background

Notation. Let $\mathbb{R}_{\geq 0} := [0, \infty)$ and $\mathbb{N} := \{0, 1, \ldots, \infty\}$. Given two vectors x and y of the same dimension, m_x denotes the dimension of x, $x^\top$ denotes the transpose of x, $|x|$ denotes the Euclidean norm of x, and $\langle x, y\rangle = x^\top y$ denotes the scalar product of x and y. Given a nonempty set $K \subset \mathbb{R}^{m_x}$, $\text{int}(K)$ denotes the interior of K, ∂K denotes its boundary, $\text{cl}(K)$ denotes its closure, and $U(K)$ denotes an open neighborhood of the set K. For a nonempty set $O \subset \mathbb{R}^{m_x}$, $K\backslash O$ denotes the subset of elements of K that are not in O. For a differentiable map $x \mapsto B(x) \in \mathbb{R}$, ∇B denotes the gradient of B with respect to x. Finally, by $\dot{x}$, we denote the time derivative of the state x, while by x^+ we denote the value of the state after an instantaneous jump.

3.1. Hybrid-model Approach

In this section, we propose a new hybrid-model approach to analyze the safety of cyber-physical systems (CPS) in the presence of attacks. Hybrid systems are models that enrich computing models with analog models of physics; as a result, they contain digital models of computing (such as automata or programs) as well as analog elements (such as differential equations) integrated in a way that allows us to reason about the effect of physics on computing and vice versa [18].

Formally, a hybrid equation is composed of a differential equation with a constraint, which models the *flow* or the continuous evolution of the system (e.g., level of water in a tank), and a difference equation with a constraint, modeling the *jumps* or discrete events (e.g., a change in the status of an actuator from ON to OFF). The strength of the hybrid equations formalism relies on the compactness of the representation and the possibility of separately using or extending the existing tools developed for continuous and discrete-time systems. Following [26], a hybrid dynamical system $\mathcal{H} = (C, F, D, G)$ as in (3) with the state variable $x \in X \subset \mathbb{R}^{m_x}$, the flow set $C \subset X$, the jump set $D \subset X$, the flow and jump maps $F : X \to X$ and $G : X \to X$, respectively.

A hybrid arc ϕ is defined on a hybrid time domain denoted $\text{dom}\,\phi \subset \mathbb{R}_{\geq 0} \times \mathbb{N}$. The hybrid arc ϕ is parametrized by an ordinary time variable $t \in \mathbb{R}_{\geq 0}$ and a discrete jump variable $j \in \mathbb{N}$. Its domain of definition $\text{dom}\,\phi$ is such that for each $(T, J) \in \text{dom}\,\phi$, $\text{dom}\,\phi \cap ([0, T] \times \{0, 1, \ldots, J\}) = \cup_{j=0}^{J} ([t_j, t_{j+1}] \times \{j\})$ for a sequence $\{t_j\}_{j=0}^{J+1}$, such that $t_{j+1} \geq t_j$, $t_0 = 0$, and $t_{j+1} = T$.

We define the concept of a *solution* x to a hybrid equation $\mathcal{H} := (C, F, D, G)$.

Definition 1 (Concept of solutions to $\mathcal{H}$). A hybrid arc $x : \text{dom}\,x \to X$ is a *solution* to $\mathcal{H}$ if

(S0) $x(0, 0) \in \text{cl}(C) \cup D$;

(S1) for all $j \in \mathbb{N}$ such that $I^j := \{t : (t, j) \in \text{dom}\,x\}$ has nonempty interior, $t \mapsto x(t, j)$ is locally absolutely continuous and

$$\begin{aligned} &x(t, j) \in C && \text{for all} \quad t \in \text{int}(I^j), \\ &\dot{x}(t, j) = F(x(t, j)) && \text{for a.a.} \quad t \in I^j; \end{aligned} \tag{1}$$

(S2) for all $(t, j) \in \text{dom}\,x$ such that $(t, j + 1) \in \text{dom}\,x$,

$$x(t, j) \in D, \qquad x(t, j + 1) = G(x(t, j)). \tag{2}$$

•

3.2. Modeling CPS using Hybrid Equations

In CPS, we identify two types of state variables. The physical variables; e.g., the water levels, which change continuously with respect to time and take values from a dense set; e.g, $\mathbb{R}_{\geq 0}$. On the other hand, the logic variables, e.g., the state of motor valves or pumps, which change through a code executed at PLCs. Those discrete variables take values from discrete sets, e.g., ON, OFF, or in transition. This heterogeneous combination of variables requires dynamical models combining continuous and discrete variables [52]. Hence, hybrid system models [26] are a natural framework for studying cyber-physical systems.

4. Proving Safety Under Attacks

4.1. Adversary Model

In the presence of actuator attacks (see Fig. 1), the attacker can falsify the actuation given to the system, either by compromising the control signal sent by the controller (right side in Fig. 1) or by compromising the actuator directly with a digital or a transduction attack [16], [54] (left side in Fig. 1). Throughout the paper, we assume the attacker can compromise one actuator and change its control action. The attacker can achieve this partial compromise by exploiting memory vulnerabilities or resource access control vulnerabilities (based on the ICS vulnerabilities categorization [60]). In this paper we focus on *post-exploitation* rather than on the specific method the attacker used to get access into the system. Our goal is to understand if a partially compromised system can remain safe.

One of our goals is to analyze security under a wide variety of attacks. In the last decade, a variety of adversary tactics have been proposed in the literature. Most of them assume that the control signal of the attacker is constrained to few parametric models. For example,

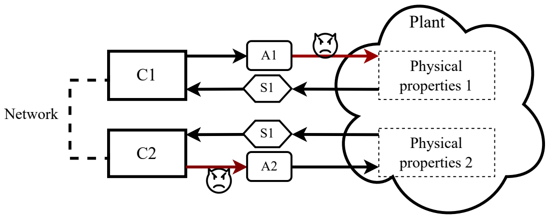


Figure 1: Simplified diagram of a CPS under actuator attacks with (C) Controllers, (S) Sensors, and (A) Actuators. Attackers can directly affect the plant via transduction attack (left side) or controller signal manipulation (right side).

a scaling attack [59] takes a compromised signal and scales it with a constant, a bias attack [12], [14] takes a compromised signal and adds a constant bias, abrupt-attacks take the maximum possible value the compromised signal can have [12], [14], [21], delay attacks take a compromised signal and delay it in time [59], and random attacks replace the compromised signal by a signal chosen from a random probability distribution [66], [20]. While all of the examples presented so far are from cyber-security conferences, the literature in control systems has very similar attack models with delay attacks [42], [32], or scaling attacks [28]. In this paper we do not place any constraints on the control signal sent by the attacker. In our case, by proposing a more general adversary model, we are not confined to existing predefined attacks.

4.2. Analyzing Safety in a CPS Under Attack

4.2.1. Safety Without Attacks. Now, we turn our attention to the concept of safety which is the property we want to analyze when the system is under attack. Intuitively safety means that the physical process will not cause harm to humans, the environment, or damage the equipment. This is best characterized by keeping a set of state variables inside a boundary (the safe set); if these variables (e.g., the level of water in a tank) go outside the safe set, then it means the system reaches an unsafe condition (i.e., the variables are in an unsafe set). To formalize this notion, we consider a hybrid system given by

$$\mathcal{H} : \begin{cases} \dot{x} = F(x) & x \in C \\ x^{+} = G(x) & x \in D, \end{cases} \tag{3}$$

and we let two sets $X_o \subset \text{cl}(C) \cup D \subset X$ and $X_u \subset X \backslash X_o$. The set X_o represents the set of initial conditions and the set X_u represents the unsafe set.

Definition 2 (Safety [47]). $\mathcal{H}$ is said to be safe with respect to (X_o, X_u) iff solutions starting at X_o never reach X_u. •

One of the main analytical tools to study safety in hybrid systems is the concept of barrier functions. A barrier function is as a scalar function of the state of the system with a given sign on the set of initial conditions X_o and the opposite sign on the unsafe set X_u.

Definition 3 (Barrier function candidate [41]). A function $B : X \to \mathbb{R}$ is a barrier function candidate with respect to (X_o, X_u) iff

$$\begin{array}{ll} B(x) > 0 & \forall x \in X_u \cap (\text{cl}(C) \cup D) \\ B(x) \leq 0 & \forall x \in X_o. \end{array} \tag{4}$$

•

Note that the barrier function candidate B in Definition 3 defines the zero-sublevel set

$$K_e := \{x \in X : B(x) \leq 0\}\,. \tag{5}$$

Notice that $X_o \subset K_e$ and $X_u \cap (C \cup D) \cap K_e = \emptyset$. Hence, safety is guaranteed provided that the barrier function candidate remains nonpositive when evaluated along the solutions starting from the initial set K_e; namely, the set K_e is forward pre-invariant. Our previous results [anonymized] identified the following sufficient conditions to certify forward pre-invariance of K_e, which in turn imply safety of $\mathcal{H}$:

$$\langle \nabla B(x), F(x) \rangle \leq 0 \qquad \forall x \in (U(\partial K_e) \backslash K_e) \cap C, \tag{6}$$

$$B(G(x)) \leq 0 \qquad \forall x \in D \cap K_e, \tag{7}$$

$$G(x) \subset C \cup D \qquad \forall x \in D \cap K_e. \tag{8}$$

To show the intuition of this approach, we illustrate how to check conditions (6)-(8) to prove the thermostat system is safe (without attacks).

Example 1 (Safety of the thermostat system). In Example 3, the objective of the control logic in (45) is to keep the physical variable z always between z_{min} and z_{max}. We now prove this claim by solving a safety problem using barrier functions. To do so, we first assume that

$$z_o < z_{min} < z_{max} < z_o + z_\Delta, \tag{9}$$

where the variables z_o and z_Δ are defined in Example 3. Furthermore, we consider the initial and unsafe sets: $X_o := \{(z, q) \in X : z \in [z_{min}, z_{max}]\}$, $X_u := X \backslash X_o$. Note that maintaining z between z_{min} and z_{max} for all time is equivalent to proving the safety of system $\mathcal{H}$ in (46) with respect to (X_o, X_u). To prove safety with respect to (X_o, X_u), we use conditions (6)-(8) and the barrier function candidate: $B(x) := (z - z_{max})(z - z_{min})$. We start noting that $C \cup D = X$ and $(1 - q, z) \in X$ for all $(q, z) \in X$. Hence, (8) is satisfied. Moreover,

$$B(G(x)) = B([(1{-}q) \quad z]^\top) = B(x) \leq 0 \quad \forall x \in K_e{\cap}D.$$

Hence, (7) is also satisfied. Finally, to verify (6), we note that, for some $\epsilon > 0$, $(U(K_e) \backslash K_e) \cap C = ((z_{max}, z_{max} + \epsilon) \times \{0\}) \cup$
$((z_{min}, z_{min} - \epsilon) \times \{1\})$. As a result, for each $x \in (U(K_e) \backslash K_e) \cap C$,

$$\langle \nabla B(x), F(x) \rangle = (z_{min}{+}z_{max}{-}2z)(z{-}z_o{-}z_\Delta q) \leq 0.$$

The latter inequality is true under (9). □

4.2.2. Safety With Attacks. As our first contribution, we now adapt our previous results to reason about safety under attacks. First, we consider a hybrid system under general attacks

$$\mathcal{H}_u : \begin{cases} \dot{x} = F(x,u) & (x,u) \in C \\ x^+ = G(x,u) & (x,u) \in D. \end{cases} \tag{10}$$

Where the attack u can affect the physical states, as well as the discrete software logic. To analyze safety in the presence of attacks, we introduce a new concept we call uniform safety.

Definition 4 (Uniform Safety). System $\mathcal{H}_u$ in (10) is said to be safe with respect to (X_o, X_u) uniformly in $u \in \mathcal{U}$ iff, for each solution pair (x,u) to $\mathcal{H}_u$ such that $x(0,0) \in X_o$, the solution x never reaches the set X_u. •

Another contribution in this paper is the derivation of new sufficient conditions to certify uniform safety of $\mathcal{H}_u$ in the presence of attacks:

$$\langle \nabla B(x), F(x,u) \rangle \leq 0 \quad \forall (x,u) \in (U(\partial K_u) \backslash K_u) \cap C, \tag{11}$$

$$B(G(x,u)) \leq 0 \quad \forall (x,u) \in D \cap K_u, \tag{12}$$

$$G(x,u) \subset C \cup D \qquad \forall (x,u) \in D \cap K_u, \tag{13}$$

where $K_u := K_e \times \mathcal{U}$, $(U(\partial K_u) \backslash K_u) = (U(\partial K_e) \backslash K_e) \times \mathcal{U}$, and $C_u := \{x \in X : \exists u \in \mathcal{U} : (x,u) \in C\}$.

Definition 5 (Barrier function certificate for safety). A $\mathcal{C}^1$ barrier function candidate with respect to (X_o, X_u) becomes a **barrier function certificate for safety** with respect to (X_o, X_u) if (6)-(8) are satisfied. •

In this paper, we extend our previous results in Theorem 5 so that we are able to prove safety in the presence of attacks.

Theorem 1. Given a hybrid system $\mathcal{H}_u = (C, F, D, G)$ as in (10), suppose that F is continuous and that there exists a $\mathcal{C}^1$ barrier function candidate B with respect to (X_o, X_u) as in (4). The hybrid system $\mathcal{H}_u$ is safe with respect to (X_o, X_u) uniformly in $u \in \mathcal{U}$ if (12) and (13) hold and

$$\langle \nabla B(x), F(x,u) \rangle \leq 0 \quad \forall (x,u) \in (U(\partial K_u) \backslash K_u) \cap C \\ : F(x,u) \in T_{C_u}(x), \tag{14}$$

where $K_u := K_e \times \mathcal{U}$, $(U(\partial K_u) \backslash K_u) = (U(\partial K_e) \backslash K_e) \times \mathcal{U}$, and $C_u := \{x \in X : \exists u \in \mathcal{U} : (x,u) \in C\}$. □

The proof of Theorem 1 can be found in Appendix B.

Fig. 11 shows that our thermostat example is unsafe under attacks, so we have to consider a more complicated toy problem (with more than one actuator) to show how to use our new results to prove safety under adversarial conditions. In the following example, we illustrate how to use conditions (11)-(13) to prove the uniform safety of a mobile robot under attacks affecting the angular-velocity actuator.

Example 2 (Proving Safety of a CPS Under Attack). Consider a robotic vehicle modeled by the kinematics equation

$$\mathcal{H}_{vu} : \begin{cases} \dot{x} = & v\cos(\theta) \\ \dot{y} = & v\sin(\theta) \\ \dot{\theta} = & u \end{cases} \qquad (x,y,\theta,v,u) \in \mathbb{R}^5, \tag{15}$$

where v and u are the forward and angular velocities, respectively. The first two elements of the state vector $[x\ y\ \theta]^\top$ correspond to the Cartesian coordinates of a point on the robot with respect to a fixed reference frame, and θ denotes the robot's orientation with respect to the same frame (see Fig. 2).

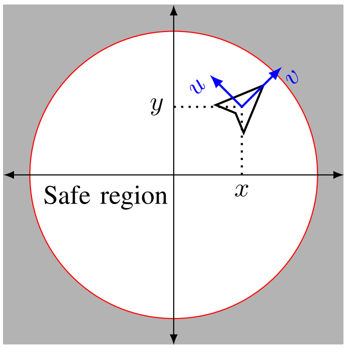


Figure 2: Robotic Vehicle and Safe Region.

The safety requirement consists in maintaining the distance between the vehicle's position and the origin within a given range. This models the case where the operator of the vehicle is at the origin and the wireless signal for operating the robot will only extend up to a radius R. If the vehicle wanders outside the safety circle, the operator will lose control of the vehicle. Hence, we assume that

$$X_u := \{(x,y,\theta) \in \mathbb{R}^3 : |(x,y)|^2 > 1\}$$
$$X_o := \{(x,y,\theta) \in \mathbb{R}^3 : |(x,y)|^2 \leq 1, \\ x\cos\theta + y\sin\theta = 0\}.$$

We now transform the coordinates of the robot from the global to the local coordinate frame; that is, we define

$$\begin{bmatrix} x_l \\ y_l \end{bmatrix} := \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{16}$$

In these new coordinates, the kinematics equation becomes

$$\mathcal{H}_u : \begin{cases} \dot{\theta} = & u \\ \dot{x}_l = & uy_l - K(x,y,\theta) \\ \dot{y}_l = & -ux_l \end{cases} \quad (x,y,\theta,u) \in \mathbb{R}^4.$$

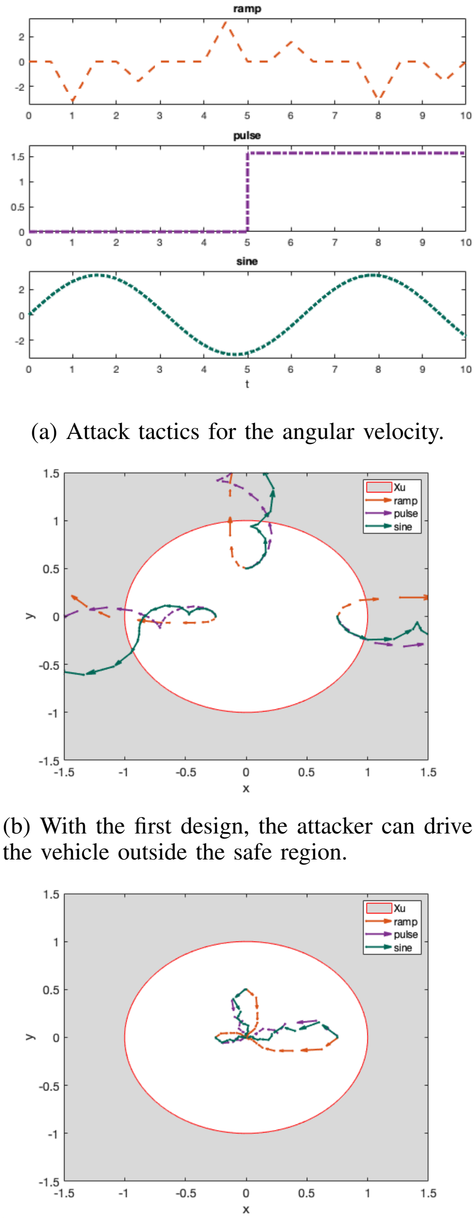
If the original control is $v := -x_l$ and $u = 0$ we show that $\mathcal{H}_u$ is safe with respect to (X_o, X_u) when there are no attacks. Indeed, for all $(x,y,\theta) \in X_o$, $F(x,y,\theta,0) = 0$. Hence, the solutions starting from X_o remain in X_o. However, as seen in Fig. 3(b), the system is not safe when u is under attack.

We now redesign v by choosing $K(x,y,\theta) := x_l$. We prove that $\mathcal{H}_u$ is safe, even under arbitrary attacks in the angular velocity. Indeed, we consider the barrier function candidate

$$B(x, y, \theta) := |(x, y)|^2 - 1 = |(x_l, y_l)|^2 - 1.$$

We now prove that $\mathcal{H}_u$ is uniformly safe with respect (X_o, X_u) by verifying (11)-(13). Indeed, note that (12) and (13) hold trivially since $\mathcal{H}_u$ is a continuous-time system. Finally, to verify (11), we note that

$$\langle \nabla B(x, y, \theta), F(x, y, \theta, u) \rangle \leq -x_l^2 \leq 0 \quad \forall (x, u) \in \mathbb{R}^4.$$



(a) Attack tactics for the angular velocity.

(b) With the first design, the attacker can drive the vehicle outside the safe region.

(c) With the second design, the system is safe under any attack on the angular velocity. Here we see responses from a ramp attack, a pulse attack and a sine attack. But the system will remain safe for any arbitrary tactic from the attacker.

Figure 3: Robotic Vehicle Example.

□

In this example, we have shown how to formally prove the safety of a vehicle when the adversary can use any arbitrary tactic to attack the angular velocity. Notice that "arbitrary" is a key concept to guarantee security. In Fig. 3(c) we see that the vehicle remains safe under three different attacks in Fig. 3(a), but how do we know the system will be safe to another attack we didn't simulate? Theorem 1 (its associated proof in Appendix B) guarantees that the system will remain safe even for attacks we have not simulated.

We now turn to the methodology introduced in this section to study SWaT.

5. A Formal Model of SWaT

In this section, we introduce a formal mathematical model of SWaT. Although SWaT has been studied extensively in security conferences [1], [2], [3], [9], [10], [12], [21], [40], [62], to the best of our knowledge, we are the first to derive and share all the equations modeling the system. We hope the new and open model in this paper will help future researchers working with SWaT.

SWaT is illustrated in Fig. 4. The control of the flow of water in the process has three stages, and each stage uses a tank to store water with different properties. Stage 1 stores raw water or pre-processing liquid, Stage 2 treats water with chemicals, and Stage 3 stores the water after filtration. The level of water in the three tanks, denoted by `L1`, `L2`, and `L3`, respectively, has to remain within a given range.

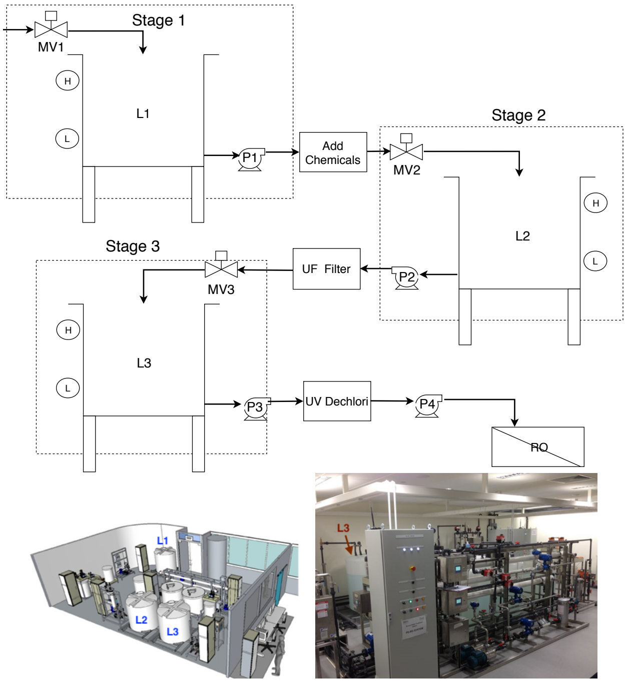


Figure 4: SWaT overview; different PLCs manage each stage.

The following components are associated with each stage:

- Motor valves `MV1`, `MV2`, and `MV3` feed tanks in each stage. Furthermore, each motor valve has four operation modes: `ON` ($\equiv 1$), `OFF` ($\equiv 0$), a transition from `ON` to `OFF` denoted T$\downarrow$ ($\equiv 3$), and a transition from `OFF` to `ON` denoted T$\uparrow$ ($\equiv 2$). Namely,

$(\mathtt{MV1},\mathtt{MV2},\mathtt{MV3}) \in \{0,1,2,3\}\times\{0,1,2,3\}\times\{0,1,2,3\}.$

- Pumps `P1`, `P2`, and `P3` drain the water from tanks to the next stage. `P1` between Stage 1 and Stage 2, `P2` between Stage 2 and Stage 3, and `P3` between Stage 3 and the final destination. Furthermore, each pump has two operation modes: it allows the water to flow when it is `ON` ($\equiv 1$), and it blocks the water flow when it is `OFF` ($\equiv 0$). Hence, $(\mathtt{P1},\mathtt{P2},\mathtt{P3}) \in \{0,1\} \times \{0,1\} \times \{0,1\}$.
- PLCs `C1`, `C2`, and `C3` control the water levels $(\mathtt{L1},\mathtt{L2},\mathtt{L3})$ in Stages 1, 2, and 3, respectively, by sending commands to $(\mathtt{MV1},\mathtt{P1})$, $(\mathtt{MV2},\mathtt{P2})$, and $(\mathtt{MV3},\mathtt{P3})$, respectively.

Strictly speaking, for each $i \in \{1,2,3\}$, the variables $(\mathtt{Pi},\mathtt{MVi})$ denote the control signals delivered by the controllers `Ci` to the $i-$th pump and the $i-$th motor valve respectively, as illustrated in Fig. 5.

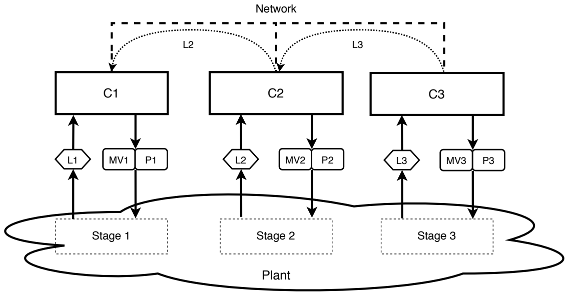


Figure 5: Computer Network of SWaT. PLCs read `L1-L3` using sensors, and control the plant through `MV` and `P`.

5.1. Discrete Dynamics

Stage 3. PLC `C3` uses `MV3` to actuate the motor valve and `P3` to actuate the pump in Stage 3. The decision to activate or deactivate the motor valve and the pump is based on the information received from a local sensor measuring the water level `L3` in the third tank.

Motor valves have four states: in addition to ON and OFF, the additional states are transition steps when migrating from 1 to 0 and vice versa. Moreover, for each $i \in \{1,2,3\}$, the transition from $\mathtt{MVi} = 3$ to $\mathtt{MVi} = 2$ cannot happen instantaneously, in the sense that the system waits T_i seconds since $\mathtt{MVi} = 3$ to switch `MVi` to 0 and then to 2. The same logic applies when transitioning from $\mathtt{MVi} = 2$ to $\mathtt{MVi} = 3$.

Each transition mode lasts for $T_3 > 0$ seconds. In summary, the control logic to actuate on the motor valve for stage 3 is

$$\mathtt{MV3} := \begin{cases} 1 & \text{if } (\tau_3 \geq T_3,\ \mathtt{MV3} = 2) \\ 3 & \text{if } (\mathtt{L3} \geq \mathtt{L3max},\ \mathtt{MV3} = 1) \\ 0 & \text{if } (\tau_3 \geq T_3,\ \mathtt{MV3} = 3) \\ 2 & \text{if } (\mathtt{L3} \leq \mathtt{L3min},\ \mathtt{MV3} = 0), \end{cases}$$

for some positive constants $\mathtt{L3max} \geq \mathtt{L3min} > 0$. Moreover, τ_3 resets to 0 each time the PLC switches the value of `MV3`.

In compact form, the discrete behavior of `MV3` and τ_3 can be modeled by the following constrained difference equation:

$$\begin{pmatrix} \tau_3^+ \\ \mathtt{MV3}^+ \end{pmatrix} = \begin{pmatrix} 0 \\ G_{\mathtt{MV3}}(\mathtt{MV3}) \end{pmatrix} \qquad (\mathtt{L3},\tau_3,\mathtt{MV3}) \in D_{\mathtt{MV3}},$$

where the set $D_{\mathtt{MV3}}$, captures the update of `MV3` and τ_3

$$\begin{aligned} D_{\mathtt{MV3}} &:= D'_{\mathtt{MV3}} \cup D''_{\mathtt{MV3}}, \\ D'_{\mathtt{MV3}} &:= \{(\mathtt{L3},\tau_3,\mathtt{MV3}) : \mathtt{L3} \leq \mathtt{L3min},\ \mathtt{MV3} = 0\} \cup \\ &\quad\ \{(\mathtt{L3},\tau_3,\mathtt{MV3}) : \mathtt{L3} \geq \mathtt{L3max},\ \mathtt{MV3} = 1\}, \\ D''_{\mathtt{MV3}} &:= \{(\mathtt{L3},\tau_3,\mathtt{MV3}) : \tau_3 \geq T_3,\ \mathtt{MV3} \in \{2,3\}\}, \end{aligned}$$

and the function $G_{\mathtt{MV3}}$, capturing the update law of `MV3`, is given by

$$G_{\mathtt{MV3}}(\mathtt{MV3}) := \begin{cases} 3 - \mathtt{MV3} & \text{if } (\mathtt{L3},\tau_3,\mathtt{MV3}) \in D''_{\mathtt{MV3}} \\ \mathtt{MV3} + 2 & \text{if } (\mathtt{L3},\tau_3,\mathtt{MV3}) \in D'_{\mathtt{MV3}}. \end{cases}$$

We also found that unless `P3` is turned off by the remote SCADA operator, it is always on.

(A1) By default, the system is constantly delivering water; namely, `P3` is always equal to 1.

Stage 2. Controller `C2` actuates the motor valve and the pump in Stage 2 using the control signals `MV2` and `P2`, respectively. Such a decision is based on the information received from a local sensor measuring the water level `L2` in Stage 2 and remote information provided by the controller `C3` concerning the level of the water `L3` in the third stage. The decision rules governing `P2` are as follows:

$$\mathtt{P2} := \begin{cases} 1 & \text{if } \ (\mathtt{L3} \leq \mathtt{L3min}, \mathtt{P2} = 0) \vee (\mathtt{MV3} \in \{1,2\}, \mathtt{P2} = 0), \\ 0 & \text{if } \ (\mathtt{L3} \geq \mathtt{L3max}, \mathtt{P2} = 1) \vee (\mathtt{MV3} \in \{0,3\}, \mathtt{P2} = 1). \end{cases} \tag{17}$$

In compact form, the behavior of `P2` can be modeled by the following constrained difference equation:

$$\mathtt{P2}^+ = G_{\mathtt{P2}}(\mathtt{P2}) \qquad (\mathtt{P2},\mathtt{L3},\mathtt{MV3}) \in D_{\mathtt{P2}},$$

where $G_{\mathtt{P2}}(\mathtt{P2}) := 1 - \mathtt{P2}$ and

$$\begin{aligned} D_{\mathtt{P2}} := &\{(\mathtt{P2},\mathtt{L3},\mathtt{MV3}) : \mathtt{L3} \leq \mathtt{L3min},\ \mathtt{P2} = 0\} \cup \\ &\{(\mathtt{P2},\mathtt{L3},\mathtt{MV3}) : \mathtt{MV3} \in \{1,2\},\ \mathtt{P2} = 0\} \cup \\ &\{(\mathtt{P2},\mathtt{L3},\mathtt{MV3}) : \mathtt{L3} \geq \mathtt{L3max},\ \mathtt{P2} = 1\} \cup \\ &\{(\mathtt{P2},\mathtt{L3},\mathtt{MV3}) : \mathtt{MV3} \in \{0,3\},\ \mathtt{P2} = 1\}. \end{aligned}$$

The motor valve in Stage 2 should go through a transition step when migrating from being 1 to 0 and vice versa. Each transition mode has a duration denoted $T_2 > 0$ seconds. To model this time delay, a timer variable $\tau_2 \in [0, T_2]$ is used. As a consequence, the decision rules governing the behavior of `MV2` are as follows:

$$\mathtt{MV2} := \begin{cases} 2 & \text{if } (\mathtt{L2} \leq \mathtt{L2min},\ \mathtt{MV2} = 0) \\ 1 & \text{if } (\tau_2 \geq T_2,\ \mathtt{MV2} = 2) \\ 3 & \text{if } (\mathtt{L2} \geq \mathtt{L2max},\ \mathtt{MV2} = 1) \\ 0 & \text{if } (\tau_2 \geq T_2,\ \mathtt{MV2} = 3), \end{cases} \tag{18}$$

for some positive constants $\mathtt{L2max} \geq \mathtt{L2min} > 0$. Moreover, we switch the value of τ_2 to 0 (timer reset) each time `MV2` switches.

In a compact form, the discrete behavior of $\texttt{MV2}$ and τ_2 can be modeled by the following constrained difference equation:

$$\begin{bmatrix} \tau_2^+ \\ \texttt{MV2}^+ \end{bmatrix} = \begin{bmatrix} 0 \\ G_{\texttt{MV2}}(\texttt{MV2}) \end{bmatrix} \qquad (\texttt{L2}, \tau_2, \texttt{MV2}) \in D_{\texttt{MV2}},$$

where $D_{\texttt{MV2}} := D'_{\texttt{MV2}} \cup D''_{\texttt{MV2}}$,

$$\begin{aligned} D'_{\texttt{MV2}} :=& \left\{(\texttt{L2}, \tau_2, \texttt{MV2}) : \texttt{L2} \geq \texttt{L2max},\ \texttt{MV2} = 1\right\} \cup \\ & \left\{(\texttt{L2}, \tau_2, \texttt{MV2}) : \texttt{L2} \leq \texttt{L2min},\ \texttt{MV2} = 0\right\}, \\ D''_{\texttt{MV2}} :=& \left\{(\texttt{L2}, \tau_2, \texttt{MV2}) : \tau_2 \geq T_2,\ \texttt{MV2} \in \{2, 3\}\right\}, \end{aligned}$$

and $G_{\texttt{MV2}}(\texttt{MV2}) := \begin{cases} 3 - \texttt{MV2} & \text{if } (\texttt{L2}, \tau_2, \texttt{MV2}) \in D''_{\texttt{MV2}} \\ \texttt{MV2} + 2 & \text{if } (\texttt{L2}, \tau_2, \texttt{MV2}) \in D'_{\texttt{MV2}}. \end{cases}$

Stage 1. Similarly to Stage 2, on Stage 1, controller $\texttt{C1}$ actuates the motor valve and the pump in Stage 1 using the control signals $\texttt{MV1}$ and $\texttt{P1}$, respectively. Such a decision is based on the information received from a local sensor measuring the water level $\texttt{L1}$ in Stage 1 and remote information provided by the controller $\texttt{C2}$ concerning the level of the water $\texttt{L2}$ in the second stage. The decision rules governing $\texttt{P1}$ are as follows:

$$\texttt{P1} := \begin{cases} 1 & \text{if} \quad (\texttt{L2} \leq \texttt{L2min}, \texttt{P1} = 0) \vee (\texttt{MV2} \in \{1, 2\}, \texttt{P1} = 0), \\ 0 & \text{if} \quad (\texttt{L2} \geq \texttt{L2max}, \texttt{P1} = 1) \vee (\texttt{MV2} \in \{0, 3\}, \texttt{P1} = 1). \end{cases} \tag{19}$$

In compact form, the behavior of $\texttt{P1}$ can be modeled by the following constrained difference equation:

$$\texttt{P1}^+ = G_{\texttt{P1}}(\texttt{P1}) \qquad (\texttt{P1}, \texttt{L2}, \texttt{MV2}) \in D_{\texttt{P1}},$$

where $G_{\texttt{P1}}(\texttt{P1}) := 1 - \texttt{P1}$ and

$$\begin{aligned} D_{\texttt{P1}} :=& \left\{(\texttt{P1}, \texttt{L2}, \texttt{MV2}) : \texttt{L2} \leq \texttt{L2min},\ \texttt{P1} = 0\right\} \cup \\ & \left\{(\texttt{P1}, \texttt{L2}, \texttt{MV2}) : \texttt{MV2} \in \{1, 2\},\ \texttt{P1} = 0\right\} \cup \\ & \left\{(\texttt{P1}, \texttt{L2}, \texttt{MV2}) : \texttt{L2} \geq \texttt{L2max},\ \texttt{P1} = 1\right\} \cup \\ & \left\{(\texttt{P1}, \texttt{L2}, \texttt{MV2}) : \texttt{MV2} \in \{0, 3\},\ \texttt{P1} = 1\right\}. \end{aligned}$$

The motor valve in Stage 1 should go through a transition step when migrating from 1 to 0 and vice versa. Each transition mode has a duration denoted $T_1 > 0$ seconds. To model this time delay, a timer variable $\tau_1 \in [0, T_1]$ is used. As a consequence, the decision rules governing the behavior of $\texttt{MV1}$ are as follows:

$$\texttt{MV1} := \begin{cases} 2 & \text{if } (\texttt{L1} \leq \texttt{L1min},\ \texttt{MV1} = 0) \\ 1 & \text{if } (\tau_1 \geq T_1,\ \texttt{MV1} = 2) \\ 3 & \text{if } (\texttt{L1} \geq \texttt{L1max},\ \texttt{MV1} = 1) \\ 0 & \text{if } (\tau_1 \geq T_1,\ \texttt{MV1} = 3), \end{cases} \tag{20}$$

for some positive constants $\texttt{L1max} \geq \texttt{L1min} > 0$. Moreover, we switch the value of τ_1 to 0 (timer reset) each time $\texttt{MV1}$ switches.

In a compact form, the discrete behavior of $\texttt{MV1}$ and τ_1 can be modeled by the following constrained difference equation:

$$\begin{bmatrix} \tau_1^+ \\ \texttt{MV1}^+ \end{bmatrix} = \begin{bmatrix} 0 \\ G_{\texttt{MV1}}(\texttt{MV1}) \end{bmatrix} \qquad (\texttt{L1}, \tau_1, \texttt{MV1}) \in D_{\texttt{MV1}},$$

where $D_{\texttt{MV1}} := D'_{\texttt{MV1}} \cup D''_{\texttt{MV1}}$,

$$\begin{aligned} D'_{\texttt{MV1}} :=& \left\{(\texttt{L1}, \tau_1, \texttt{MV1}) : \texttt{L1} \geq \texttt{L1max},\ \texttt{MV1} = 1\right\} \cup \\ & \left\{(\texttt{L1}, \tau_1, \texttt{MV1}) : \texttt{L1} \leq \texttt{L1min},\ \texttt{MV1} = 0\right\}, \\ D''_{\texttt{MV1}} :=& \left\{(\texttt{L1}, \tau_1, \texttt{MV1}) : \tau_1 \geq T_1,\ \texttt{MV1} \in \{2, 3\}\right\}, \end{aligned}$$

and $G_{\texttt{MV1}}(\texttt{MV1}) := \begin{cases} 3 - \texttt{MV1} & \text{if } (\texttt{L1}, \tau_1, \texttt{MV1}) \in D''_{\texttt{MV1}} \\ \texttt{MV1} + 2 & \text{if } (\texttt{L1}, \tau_1, \texttt{MV1}) \in D'_{\texttt{MV1}}. \end{cases}$

While we extracted the discrete model directly, we also developed an automated tool to help us with this analysis. Details are in Appendix F.

5.2. Continuous Dynamics

We now turn our attention to the continuous dynamics. In the absence of attacks, the rate of change of the water level $\texttt{L3}$ in Stage 3 depends only on the values of $\texttt{MV3}$ and $\texttt{P3}$, which coincides with the actual states of the motor valve and the pump in Stage 3:

$$\dot{\texttt{L3}} = F_{\texttt{L3}}(\texttt{MV3}, \texttt{P3}), \tag{21}$$

for some $F_{\texttt{L3}} : \{0, 1, 2, 3\} \times \{0, 1\} \to \mathbb{R}$ satisfying the following properties:

$$F_{\texttt{L3}}(2, 1) = F_{\texttt{L3}}(3, 1) > 0, \quad F_{\texttt{L3}}(0, 1) < 0, \quad F_{\texttt{L3}}(1, 1) > 0. \tag{22}$$

Similarly, the rate of change of the water level $\texttt{L2}$ in Stage 2 satisfies

$$\dot{\texttt{L2}} = F_{\texttt{L2}}(\texttt{MV2}, \texttt{P2}), \tag{23}$$

for some $F_{\texttt{L2}} : \{0, 1, 2, 3\} \times \{0, 1\} \to \mathbb{R}$ satisfying

$$\begin{gathered} F_{\texttt{L2}}(2, 0) = F_{\texttt{L2}}(3, 0) > 0, \quad F_{\texttt{L2}}(2, 1) = F_{\texttt{L2}}(3, 1) < 0, \\ F_{\texttt{L2}}(1, \texttt{P2}) > 0 \quad \forall \texttt{P2} \in \{0, 1\}, \quad F_{\texttt{L2}}(0, 1) < 0, \quad F_{\texttt{L2}}(0, 0) = 0. \end{gathered} \tag{24}$$

Furthermore, since we are not considering attacks in this section, the dynamics of $\texttt{L2}$ do not depend on $\texttt{MV3}$, which means that when the pump $\texttt{P2} = 1$ and the motor valve $\texttt{MV3} = 2$ or 1, the same water stream is removed from the second tank due to the length of the channel between $\texttt{P2}$ and $\texttt{MV3}$. Furthermore, according to (17), for each $i \in \{1, 2\}$, the pump $\texttt{Pi}$ and the motor valve $\texttt{MVi+1}$ are such that

$$\begin{aligned} \texttt{Pi} = 1 &\Longleftrightarrow \texttt{MVi+1} \in \{1, 2\} \\ \texttt{Pi} = 0 &\Longleftrightarrow \texttt{MVi+1} \in \{0, 3\}. \end{aligned} \tag{25}$$

This property implies that the stages are cascaded: the behavior of Stage 1 depends only on the variables of Stages 1 and 2, the behavior of Stage 2 depends only on the variables of Stages 2 and 3, and the behavior of Stage 3 depends only on its own variables. In addition to simplifying the analysis, (25) guarantees a safe operation mode for $\texttt{P1}$ and $\texttt{P2}$, otherwise when $\texttt{MV2} = 0$ and $\texttt{P1} = 1$, the motor valve $\texttt{MV2}$ can be damaged due to the pressure that $\texttt{P1}$ imposes.

5.3. Hybrid Model (Discrete+Continuous)

Stage 3. The state vector of Stage 3 is $x_3 := [\texttt{L3}\ \ \tau_3\ \ \texttt{MV3}]^\top \in X_3$ where $X_3 := \mathbb{R}_{\geq 0} \times [0, T_3] \times \{0, 1, 2, 3\}$. Furthermore, the hybrid system $\mathcal{H}_3$ modeling Stage 3 is given by

$$\mathcal{H}_3 : \begin{cases} \dot{x}_3 = F_3(x_3) & x_3 \in C_3 \\ x_3^+ = G_3(x_3) & x_3 \in D_3, \end{cases} \tag{26}$$

where $C_3 := \text{cl}(X_3 \backslash D_3)$, $D_3 := D_{\texttt{MV3}}$, $G_3(x_3) := [\texttt{L3}\ \ 0\ \ G_{\texttt{MV3}}(\texttt{MV3})]^\top$, $F_3(x_3) := [F_{\texttt{L3}}(\texttt{MV3}, 1)\ 1\ 0]^\top$.

Stage 2. The state vector of Stage 2 is $x_2 := [\mathtt{L2}\ \tau_2\ \mathtt{MV2}\ \mathtt{P2}]^\top \in X_2$, where $X_2 := \mathbb{R}_{\geq 0} \times [0, T_2] \times \{0,1,2,3\} \times \{0,1\}$. Furthermore, the behavior of Stage 2 is influenced by variables $(\mathtt{L3}, \mathtt{MV3})$ of Stage 3. Hence, we introduce the disturbance vector $u_2 := (\mathtt{L3}, \mathtt{MV3}) \in \mathcal{U}_2 := [\mathtt{L3min}, \mathtt{L3max} + \delta] \times \{0,1,2,3\}$, for some $\delta > 0$ to be specified later. As a result, Stage 3 can be modeled by the disturbed hybrid system $\mathcal{H}_2$ given by:

$$\mathcal{H}_2 : \begin{cases} \dot{x}_2 = F_2(x_2) & (x_2, u_2) \in C_2 \\ x_2^+ = G_2(x_2, u_2) & (x_2, u_2) \in D_2, \end{cases} \tag{27}$$

where $C_2 := \mathrm{cl}((X_2 \times \mathcal{U}_2) \backslash D_2)$, $D_2 := (\mathbb{R}_{\geq 0} \times [0, T_2] \times \{0,1,2,3\} \times D_{\mathtt{P2}}) \cup (D_{\mathtt{MV2}} \times \{0,1\} \times \mathcal{U}_2)$,

$$G_2(x_2, u_2) := [\mathtt{L2} \quad G_{22}(x_2, u_2) \quad G_{23}(x_2, u_2) \quad G_{24}(x_2, u_2)]^\top,$$

$$G_{22}(x_2, u_2) := \begin{cases} 0 & \text{if } (\mathtt{L2}, \tau_2, \mathtt{MV2}) \in D_{\mathtt{MV2}} \\ \tau_2 & \text{otherwise,} \end{cases}$$

$$G_{23}(x_2, u_2) := \begin{cases} G_{\mathtt{MV2}}(\mathtt{MV2}) & \text{if } (\mathtt{L2}, \tau_2, \mathtt{MV2}) \in D_{\mathtt{MV2}} \\ \mathtt{MV2} & \text{otherwise,} \end{cases}$$

$$G_{24}(x_2, u_2) := \begin{cases} G_{\mathtt{P2}}(\mathtt{P2}) & \text{if } (\mathtt{P2}, u_2) \in D_{\mathtt{P2}} \\ \mathtt{P2} & \text{otherwise,} \end{cases}$$

$$F_2(x_2) := [F_{\mathtt{L2}}(\mathtt{MV2}, \mathtt{P2}) \quad 1 \quad 0 \quad 0]^\top.$$

Stage 1. The hybrid equation $\mathcal{H}_1 := (C_1, F_1, D_1, G_1)$ modeling the first stage is given by:

$$\mathcal{H}_1 : \begin{cases} \dot{x}_1 = F_1(x_1) & (x_1, u_1) \in C_1 \\ x_1^+ = G_1(x_1, u_1) & (x_1, u_1) \in D_1, \end{cases} \tag{28}$$

where $x_1 := [\mathtt{L1}\ \ \tau_1\ \ \mathtt{MV1}\ \ \mathtt{P1}]^\top \in X_1 := \mathbb{R}_{\geq 0} \times [0, T_1] \times \{0,1,2,3\} \times \{0,1\}$ is the state vector of Stage 1, and $u_1 := (\mathtt{L2}, \mathtt{MV2}) \in \mathcal{U}_1 := [\mathtt{L2min} - \delta, \mathtt{L2max} + \delta] \times \{0,1,2,3\}$ is the disturbance vector formed by variables from Stage 1. Finally, the data (C_1, F_1, D_1, G_1) is constructed the same way as (C_2, F_2, D_2, G_2).

Remark 1. The behavior of the process is fully represented by the state vector $x := (x_1, x_2, x_3)$ (and associated numerical values in Appendix G) where $x_3 := (\mathtt{L3}, \tau_3, \mathtt{MV3}) \in X_3$, $x_2 := (\mathtt{L2}, \tau_2, \mathtt{MV2}, \mathtt{P2}) \in X_2$, $x_1 := (\mathtt{L1}, \tau_1, \mathtt{MV1}, \mathtt{P1}) \in X_1$. This is the first accurate and formal model of an industrial system that has been used in the past in various security conferences [62], [12], [21], [9], [3]. We believe this white-box model (in the sense that everything can be explained from first principles, as opposed to a black box model producing outputs from inputs without any explanation of their relationship) can help future researchers extend and improve their security studies for SWaT. •

Remark 2. In addition to creating the most comprehensive formal description of this popular process, we are also the first to prove that the system is safe without attacks. To analyze the safety of the water treatment plant, we consider a solution $x := (x_1, x_2, x_3)$ starting from the initial set

$$X_o := X_{o1} \times X_{o2} \times X_{o3}, \tag{29}$$

$$X_{o1} := \{x_1 \in X_1 : \mathtt{L1} \in [\mathtt{L1min}, \mathtt{L1max}]\}, \tag{30}$$

$$X_{o2} := \{x_2 \in X_2 : \mathtt{L2} \in [\mathtt{L2min}, \mathtt{L2max}]\}, \tag{31}$$

$$X_{o3} := \{x_3 \in X_3 : \mathtt{L3} \in [\mathtt{L3min}, \mathtt{L3max}]\}. \tag{32}$$

It is important to note that, due to the transition modes delaying the reaction of the motor valves, it is not possible to guarantee that such a solution x remains in the set X_o. However, we will be able to show that such a solution x remains in a larger set

$$X_s := X_{s1} \times X_{s2} \times X_{s3}, \tag{33}$$

where, for some $\delta > 0$ to be quantified, we have

$$X_{s1} := \{x_1 \in X_1 : \mathtt{L1} \in [\mathtt{L1min} - \delta, \mathtt{L1max} + \delta]\}, \tag{34}$$

$$X_{s2} := \{x_2 \in X_2 : \mathtt{L2} \in [\mathtt{L2min} - \delta, \mathtt{L2max} + \delta]\}, \tag{35}$$

$$X_{s3} := \{x_3 \in X_3 : \mathtt{L3} \in [\mathtt{L3min}, \mathtt{L3max} + \delta]\}. \tag{36}$$

As a consequence, for any unsafe set X_u satisfying $X_u \subset (X_1 \backslash X_{s1}) \times (X_2 \backslash X_{s2}) \times (X_3 \backslash X_{s3})$, we show that the plant is safe with respect to (X_o, X_u). •

Our detailed proof of safety in the absence of attacks can be found in Appendix C. From the lengthy proof in the Appendix, it is clear that these mathematical arguments are non-trivial. As far as we are aware, this is the first time this popular water treatment process has been proven safe (without attacks).

6. Security Proofs Under Actuation Attacks

Modeling actuator attacks in SWaT is not easy nor simple. Therefore, the details of the full model of SWaT under actuation attacks are presented in Appendix D. Using this attack model, we now analyze the effect of cyber-attacks and study if our new proposed countermeasures can make the system more resilient. In this section, we sketch our results, but the full proofs of adversarial safety can be found in Appendix E.

In the presence of attacks, the variables $(\mathtt{Pi}, \mathtt{MVi})$ do not necessarily correspond to the actual states of the $i-$th motor valve and the $i-$th pump, respectively. For this reason, we introduce the extra variables $(\mathtt{MV1}^a, \mathtt{MV2}^a, \mathtt{MV3}^a)$ to denote the actual states of the motor valves, $(\tau_1^a, \tau_2^a, \tau_3^a)$ to time the actual transitions of the motor valves, and $(\mathtt{P1}^a, \mathtt{P2}^a, \mathtt{P3}^a)$ to denote the actual states of the pumps.

As we will show in the next section (experimental results), the system with the original PLC programs is unsafe in the presence of attacks. Indeed, due to the constant demand of water by Stage 3, $\mathtt{L2}$ becomes less than $\mathtt{L2min} - \delta$ if $\mathtt{P1}^a = 0$ is maintained by the attacker. Similarly, $\mathtt{L3}$ becomes less than $\mathtt{L3min}$ if $\mathtt{P2}^a = 0$ is maintained. Therefore the original system is unsafe to attacks that can compromise either the first or the second pump. However, as we will show in this section, if we change the control logic of PLCs, the system can be made safe against arbitrary attacks (as long as they compromise only one control signal).

Theorem 2. Consider the hybrid system $\mathcal{H}_2$. Consider the initial set X_{o2} in (31) and the unsafe set $X_{u2} \subset$

$X_2\backslash X_{s2}$ with X_{s2} introduced in (35). Assume that there exist $\sigma_h > 0$ and $\sigma_g > 0$ such that

$$4T_2(F_{\mathrm{L2}}(2,0,0)+\sigma_h) \le \delta, \tag{37}$$
$$4T_2(F_{\mathrm{L2}}(2,1,1)+\sigma_g) \le \delta. \tag{38}$$

Then, the hybrid system $\mathcal{H}_2$ is safe with respect to (X_{o2}, X_{u2}) uniformly in $(u_2, w_2) \in \mathcal{U}_2 \times \mathcal{W}_2$, and admits a barrier function certificate given by

$$B(x_2) :=(\mathtt{L2} - \mathtt{L2min} + g(\tau_2,\mathtt{MV2}))\times$$
$$(\mathtt{L2} - \mathtt{L2max} - h(\tau_2,\mathtt{MV2})),$$

where $g(\tau_2,\mathtt{MV2}) := (-F_{\mathrm{L2}}(2,1,1)+\sigma_g)*[\tau_2 + T_2 * w_g(\mathtt{MV2})]$,
$w_g(3) := 0,\ w_g(0) := 1,\ w_g(2) := 2,\ w_g(1) := 3$,
$h(\tau_2,\mathtt{MV2}) := (F_{\mathrm{L2}}(2,0,0)+\sigma_h)*[\tau_2+T_2*w_h(\mathtt{MV2})]$,
$w_h(2) := 0,\ w_h(1) := 1,\ w_h(3) := 2,\ w_h(0) := 3$.
□

We now harden the system to make it more resilient to attacks. In particular, we first change the control logic of the PLC controlling stage 3 (C3) so that P3 is not always 1. As a result, we include P3 as a control parameter governed by the following logic:

$$\mathtt{P3} := \begin{cases} 0 & \text{if } (\mathtt{L3} \le \mathtt{L3o},\ \mathtt{P3} = 1) \\ 1 & \text{if } (\mathtt{L3} \ge \mathtt{L3o},\ \mathtt{P3} = 0), \end{cases} \tag{39}$$

where $\mathtt{L3o} > 0$ is a lower bound on the water level L3 in Stage 3, it aims to avoid the dry-runs (operates without liquid) of the pump P3. Hence, the behavior of P3 can be modeled by the following constrained difference equation:

$$\mathtt{P3}^+ = G_{\mathtt{P3}}(\mathtt{P3}) \qquad (\mathtt{L3},\mathtt{P3}) \in D_{\mathtt{P3}},$$

where $D_{\mathtt{P3}} := \{(\mathtt{L3},\mathtt{P3}) : \mathtt{L3} \le \mathtt{L3o},\ \mathtt{P3} = 1\} \cup \{(\mathtt{L3},\mathtt{P3}) : \mathtt{L3} \ge \mathtt{L3o},\ \mathtt{P3} = 0\}$ and $G_{\mathtt{P3}}(\mathtt{P3}) := 1{-}\mathtt{P3}$.

Using the logic (88) with $\mathtt{L3o} = \mathtt{L3min}$, we are able to show the following claim.

Claim 1. When P3 is governed by (88) with $\mathtt{L3o} = \mathtt{L3min}$, the plant remains safe under any arbitrary time series of possible attacks affecting $\mathtt{P2}^a$. •

To show Claim 5, we use Theorem 11 to conclude that it is enough to show the safety of Stage 3 uniformly in $\mathtt{P2}^a \in \{0,1\}$ when P3 is governed by (88). To simplify the analysis, we model Stage 3 when only $\mathtt{P2}^a$ is attacked and P3 is controlled by our modified control logic (it is hardened); see system $\mathcal{H}_3$.

Theorem 3. Consider the hybrid system $\mathcal{H}_3$. Consider the initial set $\bar{X}_{o3} := X_{o3} \times \{0,1\}$ and an unsafe set $X_{u3} \subset \bar{X}_3\backslash\bar{X}_{s3}$ with $\bar{X}_{s3} = X_{s3} \times \{0,1\}$. Assume that there exits $\sigma > 0$ such that

$$4T_3(F_{\mathrm{L3}}(3,0,1)+\sigma) \le \delta. \tag{40}$$

Then, the hybrid system $\mathcal{H}_3$ is safe with respect to $(\bar{X}_{o3}, \bar{X}_{u3})$ uniformly in $u_3 = w_{2m} = \mathtt{P2}^a \in \mathcal{U}_3$, and admits a barrier function certificate given by

$$B(\bar{x}_3) :=$$
$$(\mathtt{L3} - \mathtt{L3min})(\mathtt{L3} - \mathtt{L3max} - \mathtt{P3} * f(\tau_3,\mathtt{MV3})),$$

where $f(\tau_3,\mathtt{MV3}) := (F_{\mathrm{L3}}(3,0,1)+\sigma)[\tau_3 + T_3 * w_f(\mathtt{MV3})]$, and $w_f(2) := 0,\ w_f(1) := 1,\ w_f(3) := 2,\ w_f(0) := 3$. □

So far, we showed that our control logic modification makes the system safe against attacks in P2, but in the next section on experimental results, we show that our change is not enough when the adversary attacks P1. As a result, we need to modify the PLC controlling stage 2 as well, i.e., C2.

Claim 2. When modifying the logic in (17) governing P2 as follows:

$$\mathtt{P2} := \begin{cases} 1 \text{ if } & (\mathtt{L3} \le \mathtt{L3min}, \mathtt{P2} = 0, \mathtt{L2} \ge \mathtt{L2min}) \vee \\ & (\mathtt{MV3} \in \{1,2\}, \mathtt{P2} = 0, \mathtt{L2} \ge \mathtt{L2min}), \\ 0 \text{ if } & (\mathtt{L3} \ge \mathtt{L3max}, \mathtt{P2} = 1) \vee \\ & (\mathtt{MV3} \in \{0,3\}, \mathtt{P2} = 1) \vee \\ & (\mathtt{L2} \le \mathtt{L2min}, \mathtt{P2} = 1), \end{cases} \tag{41}$$

the plant becomes safe under any attack affecting $\mathtt{P1}^a$. •

Since only $\mathtt{P1}^a$ is attacked, the model of Stage 3 is as in (26), and its safety is already analyzed in Theorem 6. Hence, to prove Claim 6, it is enough to prove that Stage 2 is safe uniformly in $(\mathtt{P1}^a, x_3) \in \{0,1\} \times X_{s3}$ when only $\mathtt{P1}^a$ is attacked and when (91) governs P2. To simplify the proof, we model Stage 2 when only $\mathtt{P1}^a$ is attacked and when the logic governing P2 is modified;

$$\mathcal{H}_2 : \begin{cases} \dot{x}_2 = F_2(x_2,u_2) & (x_2,u_2) \in \tilde{C}_2 \times \{0,1\} \\ x_2^+ = \tilde{G}_2(x_2,\mathtt{L3},\mathtt{MV3}) & (x_2,u_2) \in \tilde{D}_2 \times \{0,1\}, \end{cases} \tag{42}$$

where

$$u_2 := (\mathtt{L3},\mathtt{MV3},\mathtt{P1}^a) \in \mathcal{U}_2,$$
$$\mathcal{U}_2 := [\mathtt{L3min}, \mathtt{L3max} + \delta] \times \{0,1,2,3\} \times \{0,1\},$$

and

$$F_2(x_2,u_2) := \begin{bmatrix} F_{\mathrm{L2}}(\mathtt{MV2},\mathtt{P2},\mathtt{P1}^a) \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

Theorem 4. Consider the hybrid system $\mathcal{H}_2$ in (92). Consider the initial set X_{o2} in (31) and the unsafe set $X_{u2} \subset X_2\backslash X_{s2}$ with X_{s2} introduced in (35). Assume that there exists $\sigma_h > 0$ such that

$$4T_2(F_{\mathrm{L2}}(2,0,1)+\sigma_h) \le \delta. \tag{43}$$

Then, the hybrid system $\mathcal{H}_2$ in (92) is safe with respect to (X_{o2}, X_{u2}) uniformly in $u_2 \in \mathcal{U}_2$, and admits a barrier function certificate given by

$$B(x_2) :=(\mathtt{L2} - \mathtt{L2min})(\mathtt{L2} - \mathtt{L2max} - \chi(\mathtt{L2}) * h(\tau_2,\mathtt{MV2})), \tag{44}$$

where $\chi : \mathbb{R} \to [0,1]$ is a smooth function such that

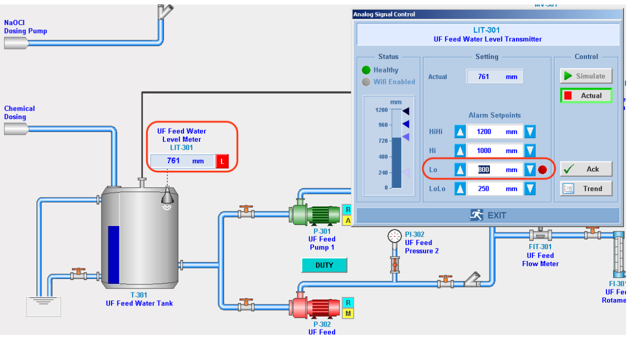
$$\begin{cases} \chi(\mathtt{L2}) = 1 & \text{if } \mathtt{L2} \ge \mathtt{L2max} \\ \chi(\mathtt{L2}) = 0 & \text{if } \mathtt{L2} \le \mathtt{L2min} \\ \chi(\mathtt{L2}) \in [0,1] & \text{otherwise,} \end{cases}$$

and $h(\tau_2,\mathtt{MV2}) := (F_{\mathrm{L2}}(2,0,1)+\sigma_h) * [\tau_2 + T_2 * w_h(\mathtt{MV2})]$, $w_h(2) := 0$, $w_h(1) := 1$, $w_h(3) := 2$, $w_h(0) := 3$. □

In summary, we have shown that the original SWaT system cannot guarantee safety when the attacker compromises any of the following actuators

P1,P2,MV1,MV2,MV3. However, we proposed a set of control logic changes to PLCs 1 and 2, and with these changes, we were able to prove that the system is safe if the attacker compromises any of these actuators: P1,P2,P3,MV2,MV3. The only time SWaT cannot guarantee safety is when the attacker compromises MV1 or compromises more than one actuator. The reason for this is that the amount of water coming into the first tank is higher than the amount that can be taken out by P1. To guarantee safety against a compromise of MV1 we would need a physical redesign of the system so that the rate of flow of entering water is the same as the rate of flow that P1 can take out of the first tank.

7. Experimental Results


Analog Signal Control

LIT-301
UF Feed Water Level Transmitter

Status: ● Healthy, ○ Wifi Enabled

Setting: Actual 761 mm

Alarm Setpoints:
HiHi 1200 mm
Hi 1000 mm
Lo 800 mm ●
LoLo 250 mm

Control: Simulate, Actual, Ack, Trend

EXIT

Figure 6: SCADA reports that the tank level is out of its operational limits; see the red box with the **L** character. In the detailed view, a red circle shows which boundary is being transgressed.

Our experiments in the real-world system confirm the theoretical results in Section 6. As depicted in Figs. 6 and 7, the original SWaT system reaches an unsafe state (L below $750mm$) showing the attacks $\mathtt{P1}^a = 0$ and $\mathtt{P2}^a = 0$ were effective.

As stated in Section E.1, PLC programs can be modified to make the system more resilient against actuator attacks. We update the PLC program, including additional validations to the control logic. The program slice in Listing 1 shows how we code Equations (91) into the PLC program. To know how to change the PLC program, we need to understand the program semantics at a high level. The PLC program uses a bit array variable (P2.SD in Listing 1) below to guarantee that the actuators always operate within the safety conditions; the code computes the safety conditions between lines 7 and 15, each one assigns to a different position in the array. Later in the code, the IF statement at line 17 checks that all safety conditions hold before turning ON the pump. To code the additional conditions, we translate Equation (91) into logical expression and insert them between lines 7 and 15.

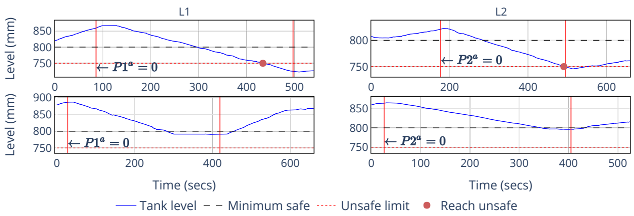


Figure 7: Response to pump attacks ($\mathtt{P1}^a$ and $\mathtt{P2}^a$). Solid red lines show when the attack starts and when it ends. Top: Tank levels reach unsafe states. Bottom: With enhanced PLC programs, controllers change the strategy when the tanks are under the lower limit.

```
1  VAR
2      P2.SD                        : ARRAY[0..15] OF BOOL;
3      P2.Auto, P2.Fault , P2.Permissive        : BOOL;
4      P2.FT_Start, P2.FT_Stop, P2. Start       : BOOL;
5  END VAR
6  ...
7  P2.SD.0 := LS2.Alarm;
8  P2.SD.1 := P2.STATUS=2 AND MV2.STATUS<>2;
9  P2.SD.2 := F2_TM.DN;
10 P2.SD.3 := L1.Level < L1min;
11 P2.SD.4 := 0;
12 ...
13 P2.SD.15:= 0;
14 ...
15 IF P2.Auto THEN
16     IF NOT P2.Fault AND NOT P2.FT_Start
17        AND NOT P2.FT_Stop AND (P2.Permissive=−1)
18        AND P2.SD=0 THEN
19           P2. Start  :=1;              (∗Turn ON P2∗)
20    ELSE IF P2.Start OR (P2.SD<>0) OR P2.Fault THEN
21          P2. Start  :=0;               (∗Turn OFF P2∗)
22     END IF;
23 END IF;
24 ...
```

Listing 1: Slice of hardened PLC program controlling P2.

We test the enhanced versions of controllers with the whole set of pump attacks. Fig. 7 (bottom) shows how the controllers mitigate the effects of the pump attacks. While the same attacks led the system to unsafe states under unprotected controllers, the enhanced version of the PLC programs allows the system to respond to the attacks effectively. Again, this matched our theoretical results in the previous section.

We now turn our attention to the fidelity of our model to the real-world system. We compare our model to traces from the real-world operation of SWaT and also to a previously proposed simulation of SWaT from a paper from IEEE S&P 2018 [12]. Fig. 8 shows how our proposed model closely follows the real-world operation of the system, while the previous simulation differs significantly from the real-world operation. In the Appendix (see Fig. 16), we include additional simulations of the proposed model to evaluate the correctness of the model empirically. We also emphasize that our implementation is based on the equations provided in this paper, while the previous simulation does not have equations for the behavior of the system, so this previous work cannot be used to reason mathematically about the safety properties of the system.

We finalize this experimental section by discussing some practicalities we discovered while trying to launch the attacks in SWaT. In our scenario, we have two pairs of Pump motor-valve in-line connections. In this setup, if the pump opens while the motor valve is closed, it will cause a dead-head effect. To mitigate the undesired effect, controllers check the motor valve state before opening the

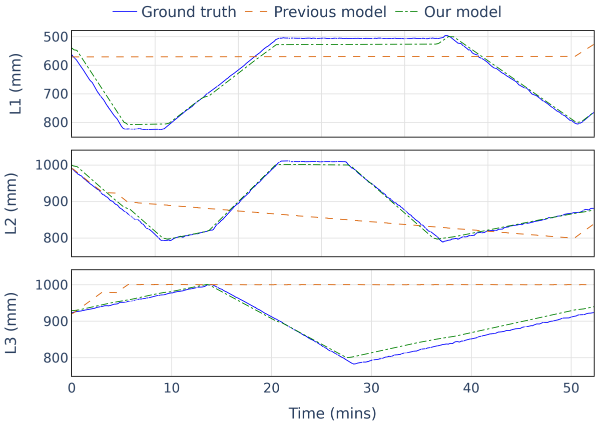
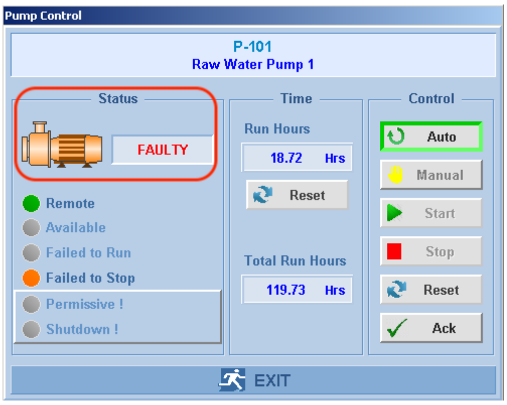


Figure 8: Comparing our SWaT model with the SWaT simulator [12] (previous model) for a period of 50 minutes.

pumps. When we launch the attack (open the pump) using the SCADA, the controller blocks the action because the next motor valve is closed. To get around this challenge, we modify the PLC code to overwrite the validation and force the output as desired. A realistic attacker can produce the same effect via a Man-in-the-middle attack, spoofing the data from the controller to the pump.

Another practical challenge we found is that after launching our attacks, the SCADA triggered an alarm linked to the pump (see Fig. 9). This is due to inconsistency. While MV is closed, the controller tries to close P, but the attacker swaps the action to open. When the SCADA reads the pump state, it reports an open state, causing the discrepancy. The attacker can hide this effect by spoofing the actuator state read back to the SCADA.


Pump Control

P-101
Raw Water Pump 1

Status: FAULTY

● Remote
● Available
● Failed to Run
● Failed to Stop
● Permissive !
● Shutdown !

Time
Run Hours: 18.72 Hrs
Total Run Hours: 119.73 Hrs

Figure 9: Effect of the attack $\texttt{P1}^a = 1$. The PLC detects pump is not OFF after sending OFF the command, so it sends this alert to the SCADA server

8. Discussion

Limitations: One limitation of our approach is that the effort in modeling real-world systems and then proving safety can be significant. Section 4 shows that analyzing two simple examples (a thermostat and a robotic vehicle) is straightforward; however, analyzing SWaT requires considerable effort. Our lengthy Appendix is an indication of the several months it took us to (1) model SWaT with hybrid mathematical equations, (2) model SWaT under actuation attacks, and (3) prove (or disprove with a counterexample) that the system is safe.

Unfortunately, there is no free lunch. Proving safety in cyber-physical systems is generally a very complex process, and state-of-the-art tools are not scalable to complex systems with hundreds or thousands of actuators (e.g., a power grid). However, several real-world control systems (like SWaT) are in the range ($\approx 20-50$ actuators and sensors) where first-principles models can be created by experts, so while our framework might not be scalable to systems with hundreds of actuators and sensors, we can still model a large amount of practical real-world control systems. Essentially, if a real-world system is similar in scale to SWaT, we can use our formal verification approaches.

Generalization: Our framework can be applied to various systems. In this paper, we have used our models to show the insecurity of a robotic vehicle and showed how to change the control algorithms to make this robotic vehicle secure against arbitrary false data injection attacks when the attacker compromises only one control signal. We then showed how to do the same analysis for a real-world water system. In general, we can use our methods in other systems where the use of barrier functions is tractable. These include bipedal robots [43], autonomous robotic systems [25], and drones [55].

Finding Barrier Functions: One of the challenges of our approach is finding barrier functions. Depending on how complex the CPS is and depending on how much understanding we have of the system's behavior, the search for barrier functions can be found intuitively by the engineer by exploiting the physics or decomposing the system into interconnected subsystems. If the system is too complex, the search for barrier functions can be performed numerically. Here, different algorithms such as SOS [47], [65] or learning-based methods [48], [49]. No matter how you look for a barrier function, your conclusions are mathematically rigorous once you find it.

9. Conclusion

In this paper, we have presented the most comprehensive formal model of a popular real-world system used by the CPS security community. We have proved for the first time that the system is safe under normal operations and also showed that the system is not safe under actuation attacks. We then showed how to modify the control logic of the PLCs so that no single actuation attack can make the system unsafe (except for MV1). This latter guarantee is proof of security against any attack tactic on any single actuator. Finally, we have shown how compromising more than one actuator results in an unsafe system.

Our main contribution is to push state-of-the-art provable security guarantees in cyber-physical systems. We argued that the progress of formal security guarantees in the past forty years has first needed precise definitions and claims. Precise definitions of security are the building block to proposing refutable assertions, allowing us to follow the scientific method because future papers can contradict or build upon them. So far, the literature on industrial control in security conferences lacks these assertions. Our paper has shown how to use the concept of barrier function certificates to develop these security assertions in a real-world process.

We hope that this work and the detailed models of this paper can help future researchers use SWaT and researchers attempting to verify CPS safety under various attacks.

References

[1] Sridhar Adepu and Aditya Mathur. Distributed detection of single-stage multipoint cyber attacks in a water treatment plant. In *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, pages 449–460, 2016.

[2] Chuadhry Mujeeb Ahmed, Martin Ochoa, Jianying Zhou, Aditya P Mathur, Rizwan Qadeer, Carlos Murguia, and Justin Ruths. Noiseprint: Attack detection using sensor and process noise fingerprint in cyber physical systems. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, pages 483–497, 2018.

[3] Chuadhry Mujeeb Ahmed, Jianying Zhou, and Aditya P Mathur. Noise matters: Using sensor and process noise fingerprint to detect stealthy cyber attacks and authenticate sensors in cps. In *Proceedings of the 34th Annual Computer Security Applications Conference*, pages 566–581, 2018.

[4] Rajeev Alur, Thao Dang, and Franjo Ivančić. Counter-example guided predicate abstraction of hybrid systems. In *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, pages 208–223. Springer, 2003.

[5] Aaron D Ames, Samuel Coogan, Magnus Egerstedt, Gennaro Notomista, Koushil Sreenath, and Paulo Tabuada. Control barrier functions: Theory and applications. In *2019 18th European control conference (ECC)*, pages 3420–3431. IEEE, 2019.

[6] AP. Revenge hacker: 34 months, must repay georgia-pacific $1m. https://www.usnews.com/news/louisiana/articles/2017-02-16/revenge-hacker-34-months-must-repay-georgia-pacific-1m, February 2017.

[7] J. P. Aubin. *Viability Theory*. Birkhauser Boston Inc., Cambridge, MA, USA, 1991.

[8] Yulong Cao, Chaowei Xiao, Benjamin Cyr, Yimeng Zhou, Won Park, Sara Rampazzi, Qi Alfred Chen, Kevin Fu, and Zhuoqing Morley Mao. Adversarial Sensor Attack on LiDAR-based Perception in Autonomous Driving. In *Conference on Computer and Communications Security (CCS)*, 2019.

[9] John H Castellanos, Martín Ochoa, and Jianying Zhou. Finding dependencies between cyber-physical domains for security testing of industrial control systems. In *Proceedings of the 34th Annual Computer Security Applications Conference*, pages 582–594, 2018.

[10] John Henry Castellanos and Jianying Zhou. A modular hybrid learning approach for black-box security testing of cps. In *International Conference on Applied Cryptography and Network Security*, pages 196–216. Springer, 2019.

[11] Xin Chen, Erika Ábrahám, and Sriram Sankaranarayanan. Flow*: An analyzer for non-linear hybrid systems. In *Computer Aided Verification: 25th International Conference, CAV 2013, Saint Petersburg, Russia, July 13-19, 2013. Proceedings 25*, pages 258–263. Springer, 2013.

[12] Yuqi Chen, Christopher M Poskitt, and Jun Sun. Learning from mutants: Using code mutation to learn and monitor invariants of a cyber-physical system. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 648–660. IEEE, 2018.

[13] Anton Cherepanov. Win32/industroyer, a new threat for industrial control systems. *White Paper. ESET*, 2017.

[14] Hongjun Choi, Wen-Chuan Lee, Yousra Aafer, Fan Fei, Zhan Tu, Xiangyu Zhang, Dongyan Xu, and Xinyan Deng. Detecting attacks against robotic vehicles: A control invariant approach. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, CCS '18, pages 801–816, New York, NY, USA, 2018. ACM.

[15] Mauro Conti, Denis Donadel, and Federico Turrin. A survey on industrial control system testbeds and datasets for security research. *IEEE Communications Surveys & Tutorials*, 23(4):2248–2294, 2021.

[16] Gökçen Yilmaz Dayanikli, Rees R Hatch, Ryan M Gerdes, Hongjie Wang, and Regan Zane. Electromagnetic sensor and actuator attacks on power converters for electric vehicles. In *2020 IEEE Security and Privacy Workshops (SPW)*, pages 98–103. IEEE, 2020.

[17] Danny Dolev and Andrew Yao. On the security of public key protocols. *IEEE Transactions on information theory*, 29(2):198–208, 1983.

[18] Laurent Doyen, Goran Frehse, George J Pappas, and André Platzer. Verification of hybrid systems. In *Handbook of Model Checking*, pages 1047–1110. Springer, 2018.

[19] Raj Gautam Dutta, Xiaolong Guo, Teng Zhang, Kevin Kwiat, Charles Kamhoua, Laurent Njilla, and Yier Jin. Estimation of safe sensor measurements of autonomous system under attack. In *Proceedings of the 54th Annual Design Automation Conference 2017*, page 46. ACM, 2017.

[20] Raj Gautam Dutta, Feng Yu, Teng Zhang, Yaodan Hu, and Yier Jin. Security for safety: a path toward building trusted autonomous vehicles. In *Proceedings of the International Conference on Computer-Aided Design*, page 92. ACM, 2018.

[21] Cheng Feng, Venkata Reddy Palleti, Aditya Mathur, and Deeph Chana. A systematic framework to generate invariants for anomaly detection in industrial control systems. In *2019 Network and Distributed System Security Symposium (NDSS)*.

[22] Xuan Feng, Qiang Li, Haining Wang, and Limin Sun. Acquisitional rule-based engine for discovering internet-of-things devices. In *27th {USENIX} Security Symposium ({USENIX} Security 18)*, pages 327–341, 2018.

[23] Ilias Giechaskiel and Kasper Rasmussen. Taxonomy and challenges of out-of-band signal injection attacks and defenses. *IEEE Communications Surveys & Tutorials*, 2019.

[24] Jairo Giraldo, Sahand Hadizadeh Kafash, Justin Ruths, and Alvaro A Cardenas. Daria: Designing actuators to resist arbitrary attacks against cyber-physical systems. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 339–353. IEEE Computer Society, 2020.

[25] Paul Glotfelter, Ian Buckley, and Magnus Egerstedt. Hybrid nonsmooth barrier functions with applications to provably safe and composable collision avoidance for robotic systems. *IEEE Robotics and Automation Letters*, 4(2):1303–1310, 2019.

[26] R. Goebel, R. G. Sanfelice, and A. R. Teel. *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton University Press, New Jersey, 2012.

[27] Oded Goldreich. *Foundations of cryptography: volume 2, basic applications*. Cambridge university press, 2009.

[28] Ziyang Guo, Dawei Shi, Karl Henrik Johansson, and Ling Shi. Optimal linear cyber-attack on remote state estimation. *IEEE Transactions on Control of Network Systems*, 4(1):4–13, 2017.

[29] Dina Hadžiosmanović, Robin Sommer, Emmanuele Zambon, and Pieter H Hartel. Through the eye of the plc: semantic security monitoring for industrial processes. In *Proceedings of the 30th Annual Computer Security Applications Conference*, pages 126–135. ACM, 2014.

[30] Changhua He, Mukund Sundararajan, Anupam Datta, Ante Derek, and John C Mitchell. A modular correctness proof of ieee 802.11 i and tls. In *Proceedings of the 12th ACM conference on Computer and communications security*, pages 2–15. ACM, 2005.

[31] Thomas A Henzinger. Hybrid automata with finite bisimulations. In *International Colloquium on Automata, Languages, and Programming*, pages 324–335. Springer, 1995.

[32] Andreas Hoehn and Ping Zhang. Detection of replay attacks in cyber-physical systems. In *2016 American Control Conference (ACC)*, pages 290–295. IEEE, 2016.

[33] Pushpak Jagtap, Sadegh Soudjani, and Majid Zamani. Temporal logic verification of stochastic systems using barrier certificates. In *International Symposium on Automated Technology for Verification and Analysis*, pages 177–193. Springer, 2018.

[34] Karl-Heinz John and Michael Tiegelkamp. *IEC 61131-3: programming industrial automation systems: concepts and programming languages, requirements for programming systems, decision-making aids*. Springer Science & Business Media, 2010.

[35] Blake Johnson, Dan Caban, Marina Krotofil, Dan Scali, Nathan Brubaker, and Christopher Glyer. Attackers deploy new ICS Attack Framework" TRITON" and cause operational disruption to critical infrastructure. *Threat Research Blog*, 2017.

[36] Sahand Hadizadeh Kafash, Jairo Giraldo, Carlos Murguia, Alvaro A Cardenas, and Justin Ruths. Constraining attacker capabilities through actuator saturation. In *2018 Annual American Control Conference (ACC)*, pages 986–991. IEEE, 2018.

[37] Jonathan Katz and Yehuda Lindell. *Introduction to modern cryptography*. CRC press, 2014.

[38] Pingfan Kong, Yi Li, Xiaohong Chen, Jun Sun, Meng Sun, and Jingyi Wang. Towards concolic testing for hybrid systems. In *International Symposium on Formal Methods*, pages 460–478. Springer, 2016.

[39] Robert M Lee, Michael J Assante, and Tim Conway. German steel mill cyber attack. *Industrial Control Systems*, 30:62, 2014.

[40] Qin Lin, Sridha Adepu, Sicco Verwer, and Aditya Mathur. Tabor: A graphical model-based approach for anomaly detection in industrial control systems. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, pages 525–536, 2018.

[41] M. Maghenem and R. G. Sanfelice. Characterizations of safety in hybrid inclusions via barrier functions. In *Proceedings of the 22Nd ACM International Conference on Hybrid Systems: Computation and Control*, HSCC '19, pages 109–118, NY, USA, 2019. ACM.

[42] Yilin Mo and Bruno Sinopoli. Secure control against replay attacks. In *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 911–918. IEEE, 2009.

[43] Quan Nguyen and Koushil Sreenath. Safety-critical control for dynamical bipedal walking with precise footstep placement. *IFAC-PapersOnLine*, 48(27):147–154, 2015.

[44] Jaclyn Peiser. A hacker broke into a florida town's water supply and tried to poison it with lye, police said. https://www.washingtonpost.com/nation/2021/02/09/oldsmar-water-supply-hack-florida/, February 2021.

[45] A. Platzer. *Logical analysis of hybrid systems: proving theorems for complex dynamics*. Springer Science & Business Media, 2010.

[46] André Platzer. The structure of differential invariants and differential cut elimination. *Logical Methods in Computer Science*, 8, 2012.

[47] S. Prajna, A. Jadbabaie, and G. J. Pappas. A framework for worst-case and stochastic safety verification using barrier certificates. *IEEE Transactions on Automatic Control*, 52(8):1415–1428, 2007.

[48] Alexander Robey, Haimin Hu, Lars Lindemann, Hanwen Zhang, Dimos V Dimarogonas, Stephen Tu, and Nikolai Matni. Learning control barrier functions from expert demonstrations. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 3717–3724. IEEE, 2020.

[49] Alexander Robey, Lars Lindemann, Stephen Tu, and Nikolai Matni. Learning robust hybrid control barrier functions for uncertain systems. *IFAC-PapersOnLine*, 54(5):1–6, 2021.

[50] Imran Sajjad, Daniel D Dunn, Rajnikant Sharma, and Ryan Gerdes. Attack mitigation in adversarial platooning using detection-based sliding mode control. In *Proceedings of the First ACM Workshop on Cyber-Physical Systems-Security and/or PrivaCy*, pages 43–53, 2015.

[51] Ricardo Sanfelice, David Copp, and Pablo Nanez. A toolbox for simulation of hybrid systems in matlab/simulink: Hybrid equations (hyeq) toolbox. In *Proceedings of the 16th international conference on Hybrid systems: computation and control*, pages 101–106, 2013.

[52] Ricardo G Sanfelice. Analysis and design of cyber-physical systems: a hybrid control systems approach. *Cyber-Physical Systems*, pages 3–31, 2015.

[53] Neetesh Saxena, Leilei Xiong, Victor Chukwuka, and Santiago Grijalva. Impact evaluation of malicious control commands in cyber-physical smart grids. *IEEE Transactions on Sustainable Computing*, 2018.

[54] Jayaprakash Selvaraj, Gökçen Y Dayanıklı, Neelam Prabhu Gaunkar, David Ware, Ryan M Gerdes, Mani Mina, et al. Electromagnetic induction attacks against embedded systems. In *Asia Conference on Computer and Communications Security (AsiaCCS)*, pages 499–510. ACM, 2018.

[55] Andrew Singletary, Aiden Swann, Yuxiao Chen, and Aaron Ames. Onboard safety guarantees for racing drones: High-speed geofencing with control barrier functions. *IEEE Robotics and Automation Letters*, 2022.

[56] Jill Slay and Michael Miller. Lessons learned from the maroochy water breach. In *Critical Infrastructure Protection*, volume 253/2007, pages 73–82. Springer Boston, November 2007.

[57] Yun Mok Son, Ho Cheol Shin, Dong Kwan Kim, Young Seok Park, Ju Hwan Noh, Ki Bum Choi, Jung Woo Choi, and Yong Dae Kim. Rocking drones with intentional sound noise on gyroscopic sensors. In *USENIX Security Symposium (USENIX Security)*. USENIX Association, 2015.

[58] Mingshun Sun, Ali Al-Hashimi, Ming Li, and Ryan Gerdes. Impacts of constrained sensing and communication based attacks on vehicular platoons. *IEEE Transactions on Vehicular Technology*, 69(5):4773–4787, 2020.

[59] Rui Tan, Varun Badrinath Krishna, David KY Yau, and Zbigniew Kalbarczyk. Impact of integrity attacks on real-time pricing in smart grids. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 439–450. ACM, 2013.

[60] Richard J Thomas and Tom Chothia. Learning from vulnerabilities-categorising, understanding and detecting weaknesses in industrial control systems. In *Computer Security: ESORICS 2020 International Workshops, CyberICPS, SECPRE, and ADIoT, Guildford, UK, September 14–18, 2020, Revised Selected Papers 6*, pages 100–116. Springer, 2020.

[61] Timothy Trippel, Ofir Weisse, Wenyuan Xu, Peter Honeyman, and Kevin Fu. Walnut: Waging doubt on the integrity of mems accelerometers with acoustic injection attacks. In *European Symposium on Security and Privacy (EuroS&P)*, pages 3–18. IEEE, 2017.

[62] David I Urbina, Jairo A Giraldo, Alvaro A Cardenas, Nils Ole Tippenhauer, Junia Valente, Mustafa Faisal, Justin Ruths, Richard Candell, and Henrik Sandberg. Limiting the impact of stealthy attacks on industrial control systems. In *Conference on Computer and Communications Security (CCS)*, pages 1092–1105. ACM, 2016.

[63] David I Urbina, Jairo Alonso Giraldo, Nils Ole Tippenhauer, and Alvaro A Cárdenas. Attacking fieldbus communications in ics: Applications to the swat testbed. In *SG-CRC*, pages 75–89, 2016.

[64] Mathy Vanhoef and Frank Piessens. Key reinstallation attacks: Forcing nonce reuse in wpa2. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1313–1328. ACM, 2017.

[65] Li Wang, Dongkun Han, and Magnus Egerstedt. Permissive barrier certificates for safe stabilization using sum-of-squares. In *2018 Annual American Control Conference (ACC)*, pages 585–590. IEEE, 2018.

[66] Yong Wang, Zhaoyan Xu, Jialong Zhang, Lei Xu, Haopei Wang, and Guofei Gu. Srid: State relation based intrusion detection for false data injection attacks in scada. In *European Symposium on Research in Computer Security*, pages 401–418. Springer, 2014.

[67] C Yan, H Shin, C Bolton, W Xu, Y Kim, and K Fu. Sok: A minimalist approach to formalizing analog sensor security. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 480–495.

[68] Kim Zetter. *Countdown to Zero Day: Stuxnet and the launch of the world's first digital weapon*. Broadway books, 2014.

[69] Lin Zhang, Xin Chen, Fanxin Kong, and Alvaro A Cardenas. Real-time attack-recovery for cyber-physical systems using linear approximations. In *2020 IEEE Real-Time Systems Symposium (RTSS)*, pages 205–217. IEEE, 2020.

A. Modeling Hybrid Systems Under Attack

As we illustrate in a canonical thermostat system, the evolution of the continuous variables is captured by a differential equation while the evolution of the discrete variables is captured by a difference equation.

Example 3 (Thermostat system model). To model a thermostat system, we use the variable $q \in \{0,1\}$, where 1 (ON) and 0 (OFF), sent by the controller to the heater. Furthermore, we use $z \in \mathbb{R}$ to denote the room's temperature. For such a system, the control logic is designed to keep the variable z between two constants z_{min} and z_{max}:

$$q := \begin{cases} 1 & \text{if } (z = z_{min},\ q = 0), \\ 0 & \text{if } (z = z_{max},\ q = 1). \end{cases} \tag{45}$$

In compact form, we can re-express (45) as the following constrained difference equation:

$$q^+ = G_q(q) \qquad (z,q) \in D_q,$$

where $G_q(q) := 1 - q$ and $D_q := \{(z,q) : z \le z_{min},\ q = 0\} \cup \{(z,q) : z \ge z_{max},\ q = 1\}$. On the other hand, the evolution of the continuous variable z is governed by the differential equation:

$$\dot{z} = F_z(z,q) := -z + z_o + z_\Delta q$$

where the parameter z_o represents the natural temperature of the room when the heater is not used and z_Δ represents the capacity of the heater to raise the temperature in the room by always being ON. As a consequence, the hybrid model of the thermostat system admits $x := (z,q) \in X := \mathbb{R} \times \{0,1\}$ as a state vector and its (hybrid) dynamics are given by

$$\mathcal{H} : \begin{cases} \dot{x} = F(x) & x \in C \\ x^+ = G(x) & x \in D, \end{cases} \tag{46}$$

where $D := D_q$, $C := \mathrm{cl}(X \backslash D)$, $G(x) := [z \quad G_q(q)]^\top$, and $F(x) := [F_z(z,q) \quad 0]^\top$. □

Thus, a single mathematical equation (46) can be used to accurately model the temperature of a room and the control actions. A simulation of this model can be seen in Fig. 10.

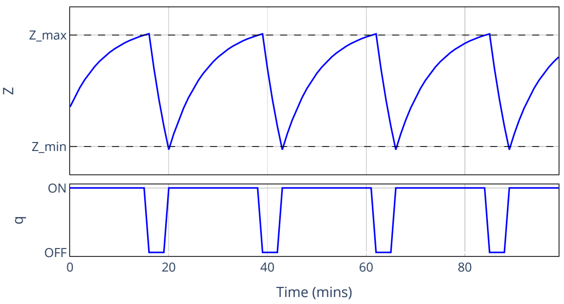


Figure 10: Flows and jumps of the thermostat system.

Example 4 (Model of the thermostat system under attack). Let $u \in \{0,1\}$ be the way the attack commands the thermostat, to turn it OFF or ON at will, so we model u as an external variable that affects the physical properties of the system and therefore we augment the original $F(x)$ with $F(x,u)$. The resulting (hybrid) dynamics are given by

$$\mathcal{H}_u : \begin{cases} \dot{x} = F(x,u) & x \in C \\ x^+ = G(x) & x \in D, \end{cases} \tag{47}$$

where $F(x,u) := [F_z(z,u) \quad 0]^\top$ and D, C, G, F_z, x are the same as in Example 3. Note that, the control variable q represents the way the system is supposed to behave, and in this case it is disregarded in $\dot{x}$ because of the attack. □

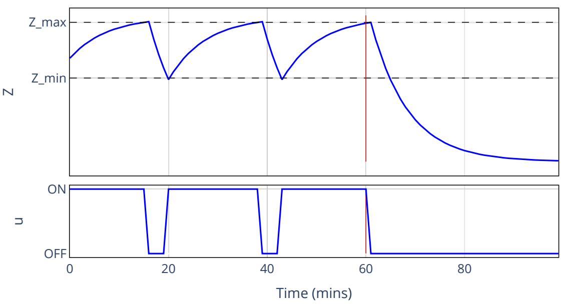


Figure 11: Attack driving the temperature of the system outside the preferred range.

Fig. 11 shows how an attack can drive the temperature outside predefined bounds.

B. New Uniform-Safety Theorems

Our previous results (Theorem 5 below) provide sufficient conditions in terms of infinitesimal inequalities — namely, without using information about solutions to the hybrid system — to guarantee that the set K_e, on which the barrier function is nonpositive, is forward invariant; namely, the solutions starting from K_e remain in K_e. More precisely, under mild conditions on the data (C, F, D, G) of the hybrid system, we present conditions for which a barrier function guarantees forward invariance of K_e. It should be noted that the proposed conditions require the barrier function to have, at points where flows are possible, a nonpositive derivative on a neighborhood of the set K_e and, after a jump from points where jumps are allowed, a nonpositive value.

Definition 6 (Contingent Cone). For a set $K \subset \mathbb{R}^{m_x}$, according to [7], the *contingent* cone of K at x is given by

$$T_K(x) := \left\{ v \in \mathbb{R}^n : \liminf_{h \to 0^+} \frac{|x + hv|_K}{h} = 0 \right\}. \tag{48}$$

•

Theorem 5. Given a hybrid system $\mathcal{H} = (C, F, D, G)$ as in (3), suppose that F is continuous and that there exists a $\mathcal{C}^1$ barrier function candidate B with respect to (X_o, X_u) as in (4). The hybrid system $\mathcal{H}$ is safe with respect to (X_o, X_u) if (7) and (8) hold and

$$\langle \nabla B(x), F(x)\rangle \leq 0 \quad \forall x \in (U(\partial K_e)\backslash K_e) \cap C\colon F(x) \in T_C(x). \tag{49}$$

□

Note that (49) is a relaxation of (6) in which we constrain the dynamics F only when it point towards the set C.

We move Theorem 1 back to the main text.

Proof. We propose to prove Theorem 1 using Theorem 5. To do so, given a exogenous signal $u : \mathbb{R}_{\geq 0} \to \mathcal{U}$, we introduce the following augmented system

$$\mathcal{H}_u : \begin{cases} \begin{bmatrix} \dot{t} \\ \dot{x} \end{bmatrix} = \begin{bmatrix} 1 \\ F(x, u(t)) \end{bmatrix} & (t, x) \in C_a \\ \begin{bmatrix} t^+ \\ x^+ \end{bmatrix} = \begin{bmatrix} t \\ G(x, u(t)) \end{bmatrix} & (t, x) \in D_a, \end{cases} \tag{50}$$

where $C_a := \{(t, x) \in \mathbb{R}_{\geq 0} \times X : (x, u(t)) \in C\}$ and $D_a := \{(t, x) \in \mathbb{R}_{\geq 0} \times X : (x, u(t)) \in D\}$. Furthermore, we consider the augmented initial and unsafe sets $X_{oa} := \mathbb{R}_{\geq 0} \times X_o$ and $X_{ou} := \mathbb{R}_{\geq 0} \times X_u$, respectively. Finally, we consider the barrier function candidate $B_a : \mathbb{R}_{\geq 0} \times X \to \mathbb{R}$ given by $B_a(t, x) := B(x)$. According to Theorem 5, the system (50) is safe with respect (X_{oa}, X_{ua}) if

$$\begin{aligned} \langle \nabla B_a(t, x), F(x, u(t))\rangle \leq 0 \quad & \forall (t, x) \in \\ & (U(\partial K_{ea})\backslash K_{ea}) \cap C_a \\ & \text{if } (1, F(x, u(t))) \in T_{C_a}(t, x), \end{aligned} \tag{51}$$

$$B_a(t, G(x, u(t))) \leq 0 \quad \forall (t, x) \in D_a \cap K_{ea}, \tag{52}$$

$$G(x, u(t)) \subset C_a \cup D_a \quad \forall (t, x) \in D_a \cap K_{ea}, \tag{53}$$

where $K_{ea} := \mathbb{R}_{\geq 0} \times K_e$. Note that when $(t, x) \in D_a \cap K_{ea}$, it follows that $(x, u(t)) \in D \cap K_u$; hence, (52) and (53) follow under (12) and (13), respectively. Furthermore, when $(t, x) \in (U(\partial K_{ea})\backslash K_{ea}) \cap C_a$, it follows that $(x, u(t)) \in (U(\partial K_u)\backslash K_u) \cap C$. Furthermore, having $(1, F(x, u(t))) \in T_{C_a}(t, x)$ implies that there exist $\{h_i\}_{i\in\mathbb{N}} \subset \mathbb{R}_{\geq 0}$ and $\{v_i\}_{i\in\mathbb{N}} \subset \mathbb{R}^{m_x+1}$, with $v_i := (v_{1i}, v_{2i}) \in \mathbb{R} \times \mathbb{R}^{m_x}$, such that $\lim_{i\to\infty} h_i \to 0^+$, $\lim_{i\to\infty} v_i = (1, F(x, u(t)))$, and $(t, x) + h_i v_i \in C_a$. Note that, for each $i \in \mathbb{N}$,

$$x + h_i v_{2i} \in \{x : \exists t \geq 0 : (t, x) \in C_a\}.$$

Hence, $x + h_i v_{2i} \in C_u$ for all $i \in \mathbb{N}$, which implies that $F(x, u(t)) \in T_{C_u}(x)$ and; thus, (51) is guaranteed under (11). ■

Definition 7 (Barrier function certificate for uniform safety). A $\mathcal{C}^1$ barrier function candidate with respect to (X_o, X_u) becomes a **barrier certificate for uniform safety** with respect to (X_o, X_u) if (11)-(13) are satisfied. •

C. Proof of Safety for SWaT Without Attacks

In this section, for each $i \in \{1, 2, 3\}$, we construct an explicit barrier certificate B guaranteeing safety of Stage i with respect to (X_{oi}, X_{ui}) uniformly in $u_i \in \mathcal{U}_i$, with $X_{ui} \subset X_i\backslash X_{si}$.

In the following result, we use Theorem 5 to show safety of $\mathcal{H}_3$ in (26).

Theorem 6. Consider the hybrid system $\mathcal{H}_3$ in (26). Consider the initial set X_{o3} in (30) and an unsafe set $X_{u3} \subset X_3\backslash X_{s3}$ with X_{s3} introduced in (36). Assume that there exits $\sigma > 0$ such that

$$4T_3(F_{\mathtt{L3}}(3) + \sigma) < \delta. \tag{54}$$

Then, the hybrid system $\mathcal{H}_3$ is safe with respect to (X_{o3}, X_{u3}), and admits the barrier function certificate

$$B(x_3) := (\mathtt{L3} - \mathtt{L3min})(\mathtt{L3} - \mathtt{L3max} - f(\tau_3, \mathtt{MV3})), \tag{55}$$

where

$$f(\tau_3, \mathtt{MV3}) := (F_{\mathtt{L3}}(3) + \sigma)[\tau_3 + T_3 * w_f(\mathtt{MV3})],$$

$$w_f(2) := 0, \; w_f(1) := 1, \; w_f(3) := 2, \; w_f(0) := 3.$$

□

Proof. Consider the hybrid system $\mathcal{H}_3$ in (26), the sets (X_{o3}, X_{u3}), and the scalar function in (55). Note that

$$f(\tau_3, \mathtt{MV3}) \in [0, \;\; (F_{\mathtt{L3}}(3) + \sigma) * 4 * T_3]$$

for all $(\tau_3, \mathtt{MV3}) \in [0, T_3] \times \{0, 1, 2, 3\}$. Hence, for $\sigma > 0$ satisfying (54), we conclude that

$$f(\tau_3, \mathtt{MV3}) \in [0, \delta]$$

for all $(\tau_3, \mathtt{MV3}) \in [0, T_3] \times \{0, 1, 2, 3\}$. Thus, (4) is satisfied. Furthermore, let the set

$$K_e := \{x_3 \in X_3 : \mathtt{L3} \in [\mathtt{L3min}, \mathtt{L3max} + f(\tau_3, \mathtt{MV3})]\}.$$

To conclude safety of the third stage using Theorem 5, we start verifying (7) and (8). To do so, we start noting that the set $K_e \cap D_3$ satisfies

$$K_e \cap D_3 = A_1 \cup A_2,$$

where

$$\begin{aligned} A_1 :=& \{x_3 \in X_3 : \mathtt{L3} \in [\mathtt{L3max}, \mathtt{L3max} + f(\tau_3, 1)], \\ & \mathtt{MV3} = \{1\}\}, \end{aligned}$$

$$\begin{aligned} A_2 :=& \{x_3 \in X_3 : \mathtt{L3} \in [\mathtt{L3min}, \mathtt{L3max} + f(T_3, \mathtt{MV3})], \\ & \tau_3 = T_3, \; \mathtt{MV3} \in \{2, 3\}\}. \end{aligned}$$

Furthermore, for each $x_3 \in A_1$, we have

$$G_3(x_3) = [\mathtt{L3} \;\, 0 \;\, 3]^\top$$

Hence,

$$B(G_3(x_3)) = (\mathtt{L3} - \mathtt{L3min})(\mathtt{L3} - \mathtt{L3max} - f(0, 3)) \leq 0.$$

The latter inequality is true since

$$f(\tau_3, 1) \leq f(0, 3) \quad \forall \tau_3 \in [0, T_3].$$

Similarly, for each $x_3 \in A_2$, we have

$$G_3(x_3) = [\mathtt{L3} \;\, 0 \;\, \alpha(\mathtt{MV3})]^\top,$$

where $\alpha(3) := 0$ and $\alpha(2) := 1$. Hence,

$$\begin{aligned} & B(G_3(x_3)) = \\ & (\mathtt{L3} - \mathtt{L3min})(\mathtt{L3} - \mathtt{L3max} - f(0, \alpha(\mathtt{MV3}))) \leq 0. \end{aligned}$$

The latter inequality is true since

$$f(\tau_3, 3) \leq f(0,0) \quad \forall \tau_3 \in [0, T_3],$$

and

$$f(\tau_3, 2) \leq f(0,1) \quad \forall \tau_3 \in [0, T_3].$$

Hence, we conclude that (7) is satisfied. Moreover, to show (8), we note that $C_3 \cup D_3 = X_3$ and $G_3(x_3) \in X_3$ for all $x_3 \in D_3$.

Next, to verify (6), we start noting that the set $U(\partial K_e) \setminus K_e$ is give by

$$U(\partial K_e) \setminus K_e = B_1 \cup B_2,$$

where, for some $\epsilon > 0$,

$$B_1 := \{x_3 \in X_3 : \texttt{L3} \in (\texttt{L3min} - \epsilon, \texttt{L3min})\},$$

and

$$B_2 := \{x_3 \in X_3 : \texttt{L3} \in (\texttt{L3max} + f(\tau_3, \texttt{MV3}))[1, \epsilon)\}.$$

Furthermore, the set C_3 can be explicitly expressed as

$$C_3 := \mathrm{cl}(X_3 \backslash D_3) = \bigcap_{i=1}^{3} \bar{D}_{3i},$$

where

$$\begin{aligned}
\bar{D}_{31} &:= \{x_3 \in X_3 : \texttt{L3} \geq \texttt{L3min} \ \cup \ \texttt{MV3} \in \{1,2,3\}\}, \\
\bar{D}_{32} &:= \{x_3 \in X_3 : \tau_3 \leq T_3 \ \cup \ \texttt{MV3} \in \{0,1\}\}, \\
\bar{D}_{33} &:= \{x_3 \in X_3 : \texttt{L3} \leq \texttt{L3max} \ \cup \ \texttt{MV3} \in \{0,2,3\}\}.
\end{aligned}$$

This is equivalent to

$$C_3 := \bigcup_{i=1}^{4} C_{3i},$$

where

$$\begin{aligned}
C_{31} &:= \{x_3 \in X_3 : \texttt{L3min} \leq \texttt{L3} \leq \texttt{L3max}\}, \\
C_{32} &:= \{x_3 \in X_3 : \texttt{L3} \leq \texttt{L3max}, \ \texttt{MV3} \in \{1,2,3\}\}, \\
C_{33} &:= \{x_3 \in X_3 : \texttt{L3} \geq \texttt{L3min}, \ \texttt{MV3} \in \{0,2,3\}\}, \\
C_{34} &:= \{x_3 \in X_3 : \texttt{MV3} \in \{2,3\}\}.
\end{aligned}$$

Hence, we conclude that

$$\begin{aligned}
&(U(\partial K_e) \setminus K_e) \cap C_3 = \\
&(B_1 \cap C_{32}) \cup (B_2 \cap C_{33}) \cup (B_1 \cap C_{34}) \cup (B_2 \cap C_{34}),
\end{aligned}$$

with

$$\begin{aligned}
&B_1 \cap C_{32} = \{x_3 \in X_3 : \\
&\texttt{L3} \in (\texttt{L3min} - \epsilon, \texttt{L3min}), \ \texttt{MV3} \in \{1,2,3\}\},
\end{aligned}$$

$$\begin{aligned}
&B_2 \cap C_{33} = \{x_3 \in X_3 : \\
&\texttt{L3} \in (\texttt{L3max} + f(\tau_3, \texttt{MV3}), \texttt{L3max} + f(\tau_3, \texttt{MV3}) + \epsilon), \\
&\texttt{MV3} \in \{0,2,3\}\},
\end{aligned}$$

$$\begin{aligned}
&B_1 \cap C_{34} = \{x_3 \in X_3 : \\
&\texttt{L3} \in (\texttt{L3min} - \epsilon, \texttt{L3min}), \ \texttt{MV3} \in \{2,3\}\},
\end{aligned}$$

$$\begin{aligned}
&B_2 \cap C_{34} = \{x_3 \in X_3 : \\
&\texttt{L3} \in (\texttt{L3max} + f(\tau_3, \texttt{MV3}), \texttt{L3max} + f(\tau_3, \texttt{MV3}) + \epsilon), \\
&\texttt{MV3} \in \{2,3\}\}.
\end{aligned}$$

Next, we evaluate the scalar product $\langle \nabla B(x_3), F_3(x_3) \rangle$ at each $x_3 \in (U(\partial K_e) \setminus K_e) \cap C_3$. Indeed, note that

$$\nabla B(x_3) = \begin{bmatrix} 2\texttt{L3} - (\texttt{L3max} + \texttt{L3min}) - f(\tau_3, \texttt{MV3}) \\ -(F_{\texttt{L3}}(3) + \sigma)(\texttt{L3} - \texttt{L3min}) \\ \star \end{bmatrix}.$$

Hence,

$$\begin{aligned}
\langle \nabla B(x_3), F(x_3) \rangle = {} & F_{\texttt{L3}}(\texttt{MV3})(\texttt{L3} - \texttt{L3max} - f(\tau_3, \texttt{MV3})) + \\
& (\texttt{L3} - \texttt{L3min})(F_{\texttt{L3}}(\texttt{MV3}) - F_{\texttt{L3}}(3) - \sigma).
\end{aligned}$$

Next, we distinguish the following two situations:

1) When $x_3 \in (B_1 \cap C_{32}) \cup (B_1 \cap C_{34})$. In this case, we conclude that $F_{\texttt{L3}}(\texttt{MV3}) \geq F_{\texttt{L3}}(3)$, $|\texttt{L3} - \texttt{L3min}| \leq \epsilon$, and $\texttt{L3max} - \texttt{L3} > \texttt{L3max} - \texttt{L3min}$. Hence,

$$\begin{aligned}
\langle \nabla B(x_3), F_3(x_3) \rangle \leq {} & F_{\texttt{L3}}(\texttt{MV3})(\texttt{L3} - \texttt{L3max}) + \\
& \epsilon |F_{\texttt{L3}}(\texttt{MV3}) - (F_{\texttt{L3}}(3) + \sigma)| \\
\leq {} & -F_{\texttt{L3}}(3) * (\texttt{L3max} - \texttt{L3min}) + \\
& \epsilon |F_{\texttt{L3}}(\texttt{MV3}) - (F_{\texttt{L3}}(3) + \sigma)|.
\end{aligned}$$

Hence, for ϵ sufficiently small, we conclude that

$$\langle \nabla B(x_3), F_3(x_3) \rangle \leq 0.$$

2) When $x_3 \in (B_2 \cap C_{33}) \cup (B_2 \cap C_{34})$. In this case, we conclude that $|\texttt{L3} - \texttt{L3max} - f(\tau_3, \texttt{MV3})| \leq \epsilon$, $\texttt{L3} - \texttt{L3min} \geq \texttt{L3max} - \texttt{L3min}$, and

$$F_{\texttt{L3}}(\texttt{MV3}) - (F_{\texttt{L3}}(3) + \sigma) \leq -\sigma.$$

Hence,

$$\begin{aligned}
\langle \nabla B(x_3), \ F_3(x_3) \rangle \leq {} & |F_{\texttt{L3}}(\texttt{MV3})| \epsilon + \\
& (\texttt{L3} - \texttt{L3min})[F_{\texttt{L3}}(\texttt{MV3}) - (F_{\texttt{L3}}(3) + \sigma)] \\
\leq {} & |F_{\texttt{L3}}(\texttt{MV3})| * \epsilon - \sigma * (\texttt{L3max} - \texttt{L3min}).
\end{aligned}$$

Hence, for ϵ sufficiently small, we conclude that

$$\langle \nabla B(x_3), F_3(x_3) \rangle \leq 0.$$

■

Next, using Theorem 1, we show safety of $\mathcal{H}_2$ in (27) uniformly in $u_2 \in \mathcal{U}_2$.

Theorem 7. Consider the hybrid system $\mathcal{H}_2$ in (27). Consider the initial set X_{o2} in (31) and the unsafe set $X_{u2} \subset X_2 \backslash X_{s2}$ with X_{s2} introduced in (35). Assume that there exit $\sigma_h > 0$ and $\sigma_g > 0$ such that

$$4T_2(F_{\mathrm{L2}}(2,0) + \sigma_h) < \delta, \tag{56}$$

$$4T_2(F_{\mathrm{L2}}(2,1) + \sigma_g) < \delta. \tag{57}$$

Then, the system $\mathcal{H}_2$ in (27) is safe with respect to (X_{o2}, X_{u2}) uniformly in $u_2 \in \mathcal{U}_2$. Moreover, the system $\mathcal{H}_2$ admits a barrier function certificate for uniform safety given by

$$\begin{aligned}
B(x_2) := {} & (\texttt{L2} - \texttt{L2min} + g(\tau_2, \texttt{MV2})) \times \\
& (\texttt{L2} - \texttt{L2max} - h(\tau_2, \texttt{MV2})),
\end{aligned} \tag{58}$$

where

$$g(\tau_2, \texttt{MV2}) := (-F_{\mathrm{L2}}(2,1) + \sigma_g) * [\tau_2 + T_2 * w_g(\texttt{MV2})],$$

$$w_g(3) := 0, \ w_g(0) := 1, \ w_g(2) := 2, \ w_g(1) := 3,$$

and

$$h(\tau_2, \texttt{MV2}) := (F_{\mathrm{L2}}(2,0) + \sigma_h) * [\tau_2 + T_2 * w_h(\texttt{MV2})],$$

$$w_h(2) := 0,\ w_h(1) := 1,\ w_h(3) := 2,\ w_h(0) := 3.$$

□

Proof. Consider the hybrid system $\mathcal{H}_2$ in (27) and the scalar function in (55). Note that, for all $(\tau_2, \mathtt{MV2}) \in [0, T_2] \times \{0, 1, 2, 3\}$, we have

$$g(\tau_2, \mathtt{MV2}) \in [0,\ ([-F_{\mathtt{L2}}(2, 1) + \sigma_g] * T_2 * 4)],$$

and

$$h(\tau_2, \mathtt{MV2}) \in [0,\ ([F_{\mathtt{L2}}(2, 0) + \sigma_h] * T_2 * 4)].$$

Hence, for σ_g and σ_h satisfying (86) and (87), we conclude that, for all $(\tau_2, \mathtt{MV2}) \in [0, T_2] \times \{0, 1, 2, 3\}$, we have

$$g(\tau_2, \mathtt{MV2}) \in [0,\ \delta] \text{ and } h(\tau_2, \mathtt{MV2}) \in [0,\ \delta].$$

Thus, (4) is satisfied. Furthermore, note that

$$K_w = \{x_2 \in X_2 :\\ \mathtt{L2} \in [\mathtt{L2min} - g(\tau_2, \mathtt{MV2}), \mathtt{L2max} + h(\tau_2, \mathtt{MV2})]\} \times \mathcal{U}_2.$$

To apply Theorem 1, we start verifying (12) and (13). To do so, we start noting that the set $K_w \cap D_2$ satisfies

$$K_w \cap D_2 = A_1 \cup A_2 \cup A_3 \cup A_4,$$

Next, to verify (11), we distinguish the following two situations:

1) When $(x_2, u_2) \in (C_{21} \cap B_1) \cup (C_{22} \cap B_1) \cup (C_{25} \cap B_1) \cup (C_{26} \cap B_1)$, we conclude that

$$\langle \nabla B(x_2), F_2(x_2) \rangle \leq \sigma_g \big[\mathtt{L2} - \mathtt{L2max} - h(\tau_2, \mathtt{MV2})\big] + \\ \epsilon[F_{\mathtt{L2}}(2, 0) + \sigma_h] \leq \\ \sigma_g \big[\mathtt{L2min} - \mathtt{L2max} - h(\tau_2, \mathtt{MV2}) - g(\tau_2, \mathtt{MV2})\big] + \\ \epsilon[F_{\mathtt{L2}}(2, 0) + \sigma_h]$$

Finally, for ϵ sufficiently small, we conclude that

$$\langle \nabla B(x_2), F_2(x_2) \rangle \leq 0.$$

2) When $(x_2, u_2) \in (C_{25} \cap B_2) \cup (C_{26} \cap B_2) \cup (C_{27} \cap B_2) \cup (C_{28} \cap B_2)$, we conclude that

$$\langle \nabla B(x_2), F_2(x_2) \rangle \leq \\ -\sigma_h[\mathtt{L2} - \mathtt{L2min} + g(\tau_2, \tau_3)] + \\ \epsilon[-F_{\mathtt{L2}}(2, 1) + \sigma_g + F_{\mathtt{L2}}(\mathtt{MV2}, \mathtt{P2})] \leq \\ -\sigma_h[\mathtt{L2max} - \mathtt{L2min} + g(\tau_2, \tau_3) + h(\tau_2, \tau_3)] + \\ \epsilon[-F_{\mathtt{L2}}(2, 1) + \sigma_g + F_{\mathtt{L2}}(\mathtt{MV2}, \mathtt{P2})].$$

Hence, for ϵ sufficiently small, we conclude that

$$\langle \nabla B(x_2), F_2(x_2) \rangle \leq 0.$$

■

The same statement as the one in Theorem 7 can be formulated for Stage 1, mutatis mutandis.

Theorem 8. Consider the hybrid system $\mathcal{H}_1$ in (28). Consider the initial set X_{o1} in (30) and the unsafe set $X_{u1} \subset X_1 \backslash X_{s1}$ with X_{s1} introduced in (34). Assume that there exit $\sigma_h > 0$ and $\sigma_g > 0$ such that

$$4T_1(F_{\mathtt{L2}}(2, 0) + \sigma_h) < \delta, \qquad (59)$$

$$4T_2(F_{\mathtt{L2}}(2, 1) + \sigma_g) < \delta. \qquad (60)$$

Then, the system $\mathcal{H}_1$ in (28) is safe with respect to (X_{o1}, X_{u1}) uniformly in $u_1 \in \mathcal{U}_1$. Moreover, the system $\mathcal{H}_1$ admits a barrier function certificate for uniform safety given by

$$B(x_1) := (\mathtt{L1} - \mathtt{L1min} + g(\tau_1, \mathtt{MV1})) \times \\ (\mathtt{L1} - \mathtt{L1max} - h(\tau_1, \mathtt{MV1})), \qquad (61)$$

where

$$g(\tau_1, \mathtt{MV1}) := (-F_{\mathtt{L1}}(2, 1) + \sigma_g) * [\tau_1 + T_1 * w_g(\mathtt{MV1})],$$

$$w_g(3) := 0,\ w_g(0) := 1,\ w_g(2) := 2,\ w_g(1) := 3,$$

and

$$h(\tau_1, \mathtt{MV1}) := (F_{\mathtt{L1}}(2, 0) + \sigma_h) * [\tau_1 + T_1 * w_h(\mathtt{MV1})],$$

$$w_h(2) := 0,\ w_h(1) := 1,\ w_h(3) := 2,\ w_h(0) := 3.$$

□

That is, the system $\mathcal{H}_1$ is safe with respect to (X_{o1}, X_{u1}) uniformly in $u_1 \in \mathcal{U}_1$ when, for some $\sigma_h > 0$ and σ_g,

$$4T_1(F_{\mathtt{L1}}(2, 0) + \sigma_h) < \delta, \qquad (62)$$

$$4T_1(F_{\mathtt{L1}}(2, 1) + \sigma_g) < \delta. \qquad (63)$$

Moreover, $\mathcal{H}_1$ admits the barrier function certificate given by

$$B(x_1) := (\mathtt{L1} - \mathtt{L1min} + g(\tau_1, \mathtt{MV1})) \times \\ (\mathtt{L1} - \mathtt{L1max} - h(\tau_1, \mathtt{MV1})). \qquad (64)$$

Finally, the combination of the previous statements allows us to conclude safety for the entire plant.

Theorem 9. Consider the hybrid system $\mathcal{H}$ composed by the cascaded interconnection of $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$. Consider the initial set X_o in (29) and an unsafe set

$$X_u \subset (X_1 \backslash X_{s1}) \times (X_2 \backslash X_{s2}) \times (X_3 \backslash X_{s3}),$$

where (X_{s1}, X_{s2}, X_{s3}) are introduced in (34), (35), and (36), respectively. Assume that there exits $\sigma > 0$, $\sigma_h > 0$, and $\sigma_g > 0$ such that (54), (86), (87), (62), and (63) hold. Then, the hybrid system $\mathcal{H}$ is safe with respect to (X_o, X_u). □

Proof. Consider the hybrid system $\mathcal{H}_2$ in (27) and the scalar function in (55). Note that, for all $(\tau_2, \mathtt{MV2}) \in [0, T_2] \times \{0, 1, 2, 3\}$, we have

$$g(\tau_2, \mathtt{MV2}) \in [0,\ ([-F_{\mathtt{L2}}(2, 1) + \sigma_g] * T_2 * 4)],$$

and

$$h(\tau_2, \mathtt{MV2}) \in [0,\ ([F_{\mathtt{L2}}(2, 0) + \sigma_h] * T_2 * 4)].$$

Hence, for σ_g and σ_h satisfying (86) and (87), we conclude that, for all $(\tau_2, \mathtt{MV2}) \in [0, T_2] \times \{0, 1, 2, 3\}$, we have

$$g(\tau_2, \mathtt{MV2}) \in [0,\ \delta] \text{ and } h(\tau_2, \mathtt{MV2}) \in [0,\ \delta].$$

Thus, (4) is satisfied. Furthermore, note that

$$K_w = \{x_2 \in X_2 :\\ \mathtt{L2} \in [\mathtt{L2min} - g(\tau_2, \mathtt{MV2}), \mathtt{L2max} + h(\tau_2, \mathtt{MV2})]\} \times \mathcal{U}_2.$$

To apply Theorem 1, we start verifying (12) and (13). To do so, we start noting that the set $K_w \cap D_2$ satisfies

$$K_w \cap D_2 = A_1 \cup A_2 \cup A_3 \cup A_4,$$

where

$$\begin{aligned} A_1 := &\{(x_2,u_2) \in X_2 \times \mathcal{U}_2 : \\ &\mathtt{L2} \in [\mathtt{L2max}, \mathtt{L2max} + h(\tau_2, 1)],\ \mathtt{MV2} = 1\}, \end{aligned}$$

$$\begin{aligned} A_2 := &\{(x_2,u_2) \in X_2 \times \mathcal{U}_2 : \\ &\mathtt{L2} \in [\mathtt{L2min} - g(\tau_2, 0), \mathtt{L2min}],\ \mathtt{MV2} = 0\}, \end{aligned}$$

$$\begin{aligned} &A_3 := \{(x_2,u_2) \in X_2 \times \mathcal{U}_2 : \\ &\mathtt{L2} \in [\mathtt{L2min} - g(T_2, \mathtt{MV2}), \mathtt{L2max} + h(T_2, \mathtt{MV2})], \\ &\tau_2 = T_2,\ \mathtt{MV2} \in \{2,3\}\}, \end{aligned}$$

$$\begin{aligned} &A_4 := \{(x_2,u_2) \in (X_2 \times \mathcal{U}_2) \backslash (A_1 \cup A_2 \cup A_3) : \\ &\mathtt{L2} \in [\mathtt{L2min} - g(\tau_2, \mathtt{MV2}), \mathtt{L2max} + h(\tau_2, \mathtt{MV2})], \\ &\mathtt{P2} = 1,\ \mathtt{L3} \geq \mathtt{L3max}\} \cup \\ &\{(x_2,u_2) \in (X_2 \times \mathcal{U}_2) \backslash (A_1 \cup A_2 \cup A_3) : \\ &\mathtt{L2} \in [\mathtt{L2min} - g(\tau_2, \mathtt{MV2}), \mathtt{L2max} + h(\tau_2, \mathtt{MV2})], \\ &\mathtt{P2} = 1,\ \mathtt{MV3} \in \{0,3\}\}, \end{aligned}$$

$$\begin{aligned} &A_5 := \{(x_2,u_2) \in (X_2 \times \mathcal{U}_2) \backslash (A_1 \cup A_2 \cup A_3) : \\ &\mathtt{L2} \in [\mathtt{L2min} - g(\tau_2, \mathtt{MV2}), \mathtt{L2max} + h(\tau_2, \mathtt{MV2})], \\ &\mathtt{P2} = 0,\ \mathtt{L3} \leq \mathtt{L3min}\} \cup \\ &\{(x_2,u_2) \in (X_2 \times \mathcal{U}_2) \backslash (A_1 \cup A_2 \cup A_3) : \\ &\mathtt{L2} \in [\mathtt{L2min} - g(\tau_2, \mathtt{MV2}), \mathtt{L2max} + h(\tau_2, \mathtt{MV2})], \\ &\mathtt{P2} = 0,\ \mathtt{MV3} \in \{1,2\}\}. \end{aligned}$$

Note that, for each $(x_2,u_2) \in A_1$, we have

$$G_2(x_2,u_2) \in [\mathtt{L2}\quad 0\quad 3\quad \{0,1\}]^\top.$$

Hence,

$$\begin{aligned} &B(G_2(x_2,u_2)) = \\ &(\mathtt{L2} - \mathtt{L2min} + g(0,3))(\mathtt{L2} - \mathtt{L2max} - h(0,3)) \leq 0. \end{aligned}$$

The latter inequality is true since $g(0,3) \geq 0$,

$$\mathtt{L2} \geq \mathtt{L2max} \quad \forall (x_2,u_2) \in A_1,$$

and

$$h(\tau_2, 1) \leq h(0,3) \quad \forall \tau_2 \in [0, T_2].$$

Similarly, for each $(x_2,u_2) \in A_2$, we have

$$G_2(x_2,u_2) = [\mathtt{L2}\ 0\ 2\ \{0,1\}]^\top.$$

Hence,

$$\begin{aligned} &B(G_2(x_2,u_2)) = \\ &(\mathtt{L2} - \mathtt{L2min} + g(0,2))(\mathtt{L2} - \mathtt{L2max} - h(0,2)) \leq 0. \end{aligned}$$

The latter inequality is true since $h(0,2) \geq 0$,

$$\mathtt{L2} \leq \mathtt{L2min} \quad \forall (x_2,u_2) \in A_2,$$

and

$$g(\tau_2, 0) \leq g(0,2) \quad \forall \tau_2 \in [0, T_2].$$

Now, for each $(x_2,u_2) \in A_3$,

$$G_2(x_2) \in [\mathtt{L2}\ 0\ \alpha(\mathtt{MV2})\ \{0,1\}]^\top,$$

where $\alpha(3) := 0$ and $\alpha(2) := 1$. Hence,

$$\begin{aligned} B(G_2(x_2)) = &(\mathtt{L2} - \mathtt{L2min} + g(0, \alpha(\mathtt{MV2}))) * \\ &(\mathtt{L2} - \mathtt{L2max} - h(0, \alpha(\mathtt{MV2}))) \leq 0. \end{aligned}$$

The latter inequality is true since

$$h(7,3) \leq h(0,0), \quad h(7,2) \leq h(0,1),$$

$$g(7,3) \leq g(0,0), \quad \text{and} \quad g(7,2) \leq g(0,1).$$

Finally, for each $(x_2,u_2) \in A_4 \cup A_5$, we have

$$G_2(x_2,u_2) \in [\mathtt{L2}\quad \tau_2\quad \mathtt{MV2}\quad \{0,1\}]^\top.$$

Hence,

$$\begin{aligned} B(G_2(x_2,u_2)) = &(\mathtt{L2} - \mathtt{L2min} + g(\tau_2, \mathtt{MV2})) * \\ &(\mathtt{L2} - \mathtt{L2max} - h(\tau_2\ \mathtt{MV2})) \leq 0. \end{aligned}$$

The latter inequality is true since $(A_4 \cup A_5) \subset K_w$. Hence, we conclude that (12) is satisfied. Moreover, to verify (13), we note that $C_2 \cup D_2 = X_2 \times \mathcal{U}_2$ and $G_2(x_2,u_2) \in X_2$ for all $(x_2,u_2) \in D_2$.

Next, to verify (11), we start noting that

$$U(\partial K_w) \setminus K_w = B_1 \cup B_2,$$

where, for some $\epsilon > 0$ sufficiently small,

$$\begin{aligned} &B_1 := \{(\mathtt{L2}, u_2) \in X_2 \times \mathcal{U}_2 : \\ &\mathtt{L2} \in (\mathtt{L2min} - g(\tau_2, \mathtt{MV2}) - \epsilon, \mathtt{L2min} - g(\tau_2, \mathtt{MV2}))\}, \end{aligned}$$

and

$$\begin{aligned} &B_2 := \{(\mathtt{L2}, u_2) \in X_2 \times \mathcal{U}_2 : \\ &\mathtt{L2} \in (\mathtt{L2max} + h(\tau_2, \mathtt{MV2}), \mathtt{L2max} + h(\tau_2, \mathtt{MV2}) + \epsilon)\}. \end{aligned}$$

Furthermore, the set C_2 can be explicitly expressed as

$$C_2 = \bigcup_{i=1}^{8} C_{2i},$$

where

$$C_{2i} := C_{2i}^a \cup C_{2i}^b \qquad \forall i \in \{2,4,6,8\},$$

and

$$\begin{aligned} C_{21} := &\big\{(x_2,u_2) \in X_2 \times \mathcal{U}_2 : \mathtt{L2} \leq \mathtt{L2max},\ \mathtt{MV2} = \{1,2,3\}, \\ &\mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\}, \end{aligned}$$

$$\begin{aligned} C_{22}^a := &\big\{(x_2,u_2) \in X_2 \times \mathcal{U}_2 : \mathtt{L2} \leq \mathtt{L2max},\ \mathtt{MV2} = \{1,2,3\}, \\ &\mathtt{L3} \leq \mathtt{L3max},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\}, \\ C_{22}^b := &\big\{(x_2,u_2) \in X_2 \times \mathcal{U}_2 : \mathtt{L2} \leq \mathtt{L2max},\ \mathtt{MV2} = \{1,2,3\}, \\ &\mathtt{L3} \leq \mathtt{L3max},\ \mathtt{P2} = \{1\},\ \mathtt{MV3} = \{1,2\}\big\}, \end{aligned}$$

$$\begin{aligned} C_{23} := &\big\{(x_2,u_2) \in X_2 \times \mathcal{U}_2 : \\ &\mathtt{L2min} \leq \mathtt{L2} \leq \mathtt{L2max},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\}, \end{aligned}$$

$$\begin{aligned} C_{24}^a := &\big\{(x_2,u_2) \in X_2 \times \mathcal{U}_2 : \mathtt{L2min} \leq \mathtt{L2} \leq \mathtt{L2max}, \\ &\mathtt{L3} \leq \mathtt{L3max},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\}, \\ C_{24}^b := &\big\{(x_2,u_2) \in X_2 \times \mathcal{U}_2 : \mathtt{L2min} \leq \mathtt{L2} \leq \mathtt{L2max}, \\ &\mathtt{L3} \leq \mathtt{L3max},\ \mathtt{P2} = \{1\},\ \mathtt{MV3} = \{1,2\}\big\}, \end{aligned}$$

$$\begin{aligned}C_{25} :=&\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 :\\ &\mathtt{MV2}\in\{2,3\},\ \mathtt{P2}=\{0\},\ \mathtt{MV3}=\{0,3\}\big\},\end{aligned}$$

$$\begin{aligned}C_{26}^{a} :=&\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 : \mathtt{MV2}\in\{2,3\},\ \mathtt{L3}\le\mathtt{L3max},\\ &\mathtt{P2}=\{0\},\ \mathtt{MV3}=\{0,3\}\big\},\\ C_{26}^{b} :=&\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 : \mathtt{MV2}\in\{2,3\},\ \mathtt{L3}\le\mathtt{L3max},\\ &\mathtt{P2}=\{1\},\ \mathtt{MV3}=\{1,2\}\big\},\end{aligned}$$

$$\begin{aligned}C_{27} :=&\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 : \mathtt{L2}\ge\mathtt{L2min},\ \mathtt{MV2}=\{0,2,3\},\\ &\mathtt{P2}=\{0\},\ \mathtt{MV3}=\{0,3\}\big\},\end{aligned}$$

$$\begin{aligned}C_{28}^{a} :=&\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 : \mathtt{L2}\ge\mathtt{L2min},\ \mathtt{MV2}=\{0,2,3\},\\ &\mathtt{L3}\in[\mathtt{L3min},\mathtt{L3max}],\ \mathtt{P2}=\{0\},\ \mathtt{MV3}=\{0,3\}\big\}.\\ C_{28}^{b} :=&\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 : \mathtt{L2}\ge\mathtt{L2min},\ \mathtt{MV2}=\{0,2,3\},\\ &\mathtt{L3}\in[\mathtt{L3min},\mathtt{L3max}],\ \mathtt{P2}=\{1\},\ \mathtt{MV3}=\{1,2\}\big\}.\end{aligned}$$

Hence, we conclude that

$$\begin{aligned}&(U(\partial K_w)\setminus K_w)\cap C_2=(C_{21}\cap B_1)\cup(C_{22}\cap B_1)\cup\\ &(C_{25}\cap B_1)\cup(C_{25}\cap B_2)\cup(C_{26}\cap B_1)\cup(C_{26}\cap B_2)\cup\\ &(C_{27}\cap B_2)\cup(C_{28}\cap B_2),\end{aligned}$$

with

$$\begin{aligned}&C_{21}\cap B_1=\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 :\\ &\mathtt{L2}\in(\mathtt{L2min}-g(\tau_2,\mathtt{MV2})-\epsilon,\mathtt{L2min}-g(\tau_2,\mathtt{MV2})),\\ &\mathtt{MV2}=\{1,2,3\},\ \mathtt{P2}=\{0\},\ \mathtt{MV3}=\{0,3\}\big\},\end{aligned}$$

$$\begin{aligned}&C_{22}\cap B_1=\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 :\\ &\mathtt{L2}\in(\mathtt{L2min}-g(\tau_2,\mathtt{MV2})-\epsilon,\mathtt{L2min}-g(\tau_2,\mathtt{MV2})),\\ &\mathtt{MV2}=\{1,2,3\},\ \mathtt{L3}\le\mathtt{L3max},\ \mathtt{P2}=\{0\},\ \mathtt{MV3}=\{0,3\}\big\}\cup\\ &\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 :\\ &\mathtt{L2}\in(\mathtt{L2min}-g(\tau_2,\mathtt{MV2})-\epsilon,\mathtt{L2min}-g(\tau_2,\mathtt{MV2})),\\ &\mathtt{MV2}=\{1,2,3\},\ \mathtt{L3}\le\mathtt{L3max},\ \mathtt{P2}=\{1\},\ \mathtt{MV3}=\{1,2\}\big\},\end{aligned}$$

$$\begin{aligned}&C_{25}\cap B_1=\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 :\\ &\mathtt{L2}\in(\mathtt{L2min}-g(\tau_2,\mathtt{MV2})-\epsilon,\mathtt{L2min}-g(\tau_2,\mathtt{MV2})),\\ &\mathtt{MV2}\in\{2,3\},\ \mathtt{P2}=\{0\},\ \mathtt{MV3}=\{0,3\}\big\},\end{aligned}$$

$$\begin{aligned}&C_{26}\cap B_1=\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 :\\ &\mathtt{L2}\in(\mathtt{L2min}-g(\tau_2,\mathtt{MV2})-\epsilon,\mathtt{L2min}-g(\tau_2,\mathtt{MV2})),\\ &\mathtt{MV2}\in\{2,3\},\ \mathtt{L3}\le\mathtt{L3max},\ \mathtt{P2}=\{0\},\ \mathtt{MV3}=\{0,3\}\big\}\cup\\ &\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 :\\ &\mathtt{L2}\in(\mathtt{L2min}-g(\tau_2,\mathtt{MV2})-\epsilon,\mathtt{L2min}-g(\tau_2,\mathtt{MV2})),\\ &\mathtt{MV2}\in\{2,3\},\ \mathtt{L3}\le\mathtt{L3max},\ \mathtt{P2}=\{1\},\ \mathtt{MV3}=\{1,2\}\big\},\end{aligned}$$

$$\begin{aligned}&C_{25}\cap B_2=\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 :\\ &\mathtt{L2}\in(\mathtt{L2max}+h(\tau_2,\mathtt{MV2}),\mathtt{L2max}+h(\tau_2,\mathtt{MV2})+\epsilon),\\ &\mathtt{MV2}\in\{2,3\},\ \mathtt{P2}=\{0\},\ \mathtt{MV3}=\{0,3\}\big\},\end{aligned}$$

$$\begin{aligned}&C_{26}\cap B_2=\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 :\\ &\mathtt{L2}\in(\mathtt{L2max}+h(\tau_2,\mathtt{MV2}),\mathtt{L2max}+h(\tau_2,\mathtt{MV2})+\epsilon),\\ &\mathtt{MV2}\in\{2,3\},\ \mathtt{L3}\le\mathtt{L3max},\ \mathtt{P2}=\{0\},\ \mathtt{MV3}=\{0,3\}\big\}\cup\\ &\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 :\\ &\mathtt{L2}\in(\mathtt{L2max}+h(\tau_2,\mathtt{MV2}),\mathtt{L2max}+h(\tau_2,\mathtt{MV2})+\epsilon),\\ &\mathtt{MV2}\in\{2,3\},\ \mathtt{L3}\le\mathtt{L3max},\ \mathtt{P2}=\{1\},\ \mathtt{MV3}=\{1,2\}\big\},\end{aligned}$$

$$\begin{aligned}&C_{27}\cap B_2=\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 :\\ &\mathtt{L2}\in(\mathtt{L2max}+h(\tau_2,\mathtt{MV2}),\mathtt{L2max}+h(\tau_2,\mathtt{MV2})+\epsilon),\\ &\mathtt{MV2}=\{0,2,3\},\ \mathtt{P2}=\{0\},\ \mathtt{MV3}=\{0,3\}\big\},\end{aligned}$$

$$\begin{aligned}&C_{28}\cap B_2=\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 :\\ &\mathtt{L2}\in(\mathtt{L2max}+h(\tau_2,\mathtt{MV2}),\mathtt{L2max}+h(\tau_2,\mathtt{MV2})+\epsilon),\\ &\mathtt{MV2}=\{0,2,3\},\ \mathtt{L3}\in[\mathtt{L3min},\mathtt{L3max}],\ \mathtt{P2}=\{1\},\\ &\mathtt{MV3}=\{1,2\}\big\}\cup\big\{(x_2,u_2)\in X_2\times\mathcal{U}_2 :\\ &\mathtt{L2}\in(\mathtt{L2max}+h(\tau_2,\mathtt{MV2}),\mathtt{L2max}+h(\tau_2,\mathtt{MV2})+\epsilon),\\ &\mathtt{MV2}=\{0,2,3\},\ \mathtt{L3}\in[\mathtt{L3min},\mathtt{L3max}],\ \mathtt{P2}=\{0\},\\ &\mathtt{MV3}=\{0,3\}\big\},\end{aligned}$$

Next, we evaluate the scalar product $\langle\nabla B(x_2),F_2(x_2)\rangle$ for each $(x_2,u_2)\in(U(\partial K_w)\setminus K_w)\cap C_2$. To do so, we note that

$$\nabla B(x_2)=[\nabla_{\mathtt{L2}}B(x_2)\ \nabla_{\tau_2}B(x_2)\ \star\ \star]^{\top},$$

where

$$\begin{aligned}\nabla_{\mathtt{L2}}B(x_2):=\ &2\mathtt{L2}-(\mathtt{L2max}+\mathtt{L2min})-h(\tau_2,\mathtt{MV2})+\\ &g(\tau_2,\mathtt{MV2}),\end{aligned}$$

$$\begin{aligned}\nabla_{\tau_2}B(x_2):=\ &-[\mathtt{L2}-\mathtt{L2min}+g(\tau_2,\tau_3)]*\\ &[F_{\mathtt{L2}}(2,0)+\sigma_h]+\\ &[\mathtt{L2}-\mathtt{L2max}-h(\tau_2,\mathtt{MV2})]*\\ &[-F_{\mathtt{L2}}(2,1)+\sigma_g].\end{aligned}$$

Hence,

$$\begin{aligned}&\langle\nabla B(x_2),F(x_2)\rangle=F_{\mathtt{L2}}(\mathtt{MV2},\mathtt{P2})*\\ &\big[2\mathtt{L2}-(\mathtt{L2min}+\mathtt{L2max})-h(\tau_2,\mathtt{MV2})+\\ &g(\tau_2,\mathtt{MV2})\big]-\\ &[\mathtt{L2}-\mathtt{L2min}+g(\tau_2,\tau_3)][F_{\mathtt{L2}}(2,0)+\sigma_h]+\\ &[\mathtt{L2}-\mathtt{L2max}-h(\tau_2,\mathtt{MV2})][-F_{\mathtt{L2}}(2,1)+\sigma_g],\end{aligned}$$

and here we distinguish the following two situations:

1) When $(x_2,u_2)\in(C_{21}\cap B_1)\cup(C_{22}\cap B_1)\cup(C_{25}\cap B_1)\cup(C_{26}\cap B_1)$, we conclude that

$$F_{\mathtt{L2}}(\mathtt{MV2},\mathtt{P2})\ge F_{\mathtt{L2}}(2,1),$$
$$-\epsilon\le\mathtt{L2}-\mathtt{L2min}+g(\tau_2,\tau_3)\le 0.$$

Hence,

$$\begin{aligned}&\langle \nabla B(x_2), F_2(x_2)\rangle \le F_{\mathrm{L2}}(\mathtt{MV2},\mathtt{P2})\times\\ &\big[\mathtt{L2}-\mathtt{L2max}-h(\tau_2,\mathtt{MV2})\big]+\\ &\epsilon[F_{\mathrm{L2}}(2,0)+\sigma_h]+\\ &[\mathtt{L2}-\mathtt{L2max}-h(\tau_2,\mathtt{MV2})][-F_{\mathrm{L2}}(2,1)+\sigma_g]\end{aligned}$$

and, thus,

$$\begin{aligned}&\langle \nabla B(x_2), F_2(x_2)\rangle \le \sigma_g\big[\mathtt{L2}-\mathtt{L2max}-h(\tau_2,\mathtt{MV2})\big]+\\ &\epsilon[F_{\mathrm{L2}}(2,0)+\sigma_h] \le\\ &\sigma_g\big[\mathtt{L2min}-\mathtt{L2max}-h(\tau_2,\mathtt{MV2})-g(\tau_2,\mathtt{MV2})\big]+\\ &\epsilon[F_{\mathrm{L2}}(2,0)+\sigma_h]\end{aligned}$$

Finally, for ϵ sufficiently small, we conclude that

$$\langle \nabla B(x_2), F_2(x_2)\rangle \le 0.$$

2) When $(x_2,u_2)\in(C_{25}\cap B_2)\cup(C_{26}\cap B_2)\cup(C_{27}\cap B_2)\cup(C_{28}\cap B_2)$, we conclude that

$$F_{\mathrm{L2}}(\mathtt{MV2},\mathtt{P2}) \le F_{\mathrm{L2}}(2,0),$$
$$0 \le \mathtt{L2}-\mathtt{L2max}-h(\tau_2,\tau_3)\le\epsilon.$$

Hence,

$$\begin{aligned}&\langle \nabla B(x_2), F_2(x_2)\rangle \le F_{\mathrm{L2}}(\mathtt{MV2},\mathtt{P2})*\\ &\big[\mathtt{L2}-\mathtt{L2min}+g(\tau_2,\mathtt{MV2})+\epsilon\big]-\\ &[\mathtt{L2}-\mathtt{L2min}+g(\tau_2,\tau_3)][F_{\mathrm{L2}}(2,0)+\sigma_h]+\\ &\epsilon[-F_{\mathrm{L2}}(2,1)+\sigma_g]\end{aligned}$$

and, thus,

$$\begin{aligned}&\langle \nabla B(x_2), F_2(x_2)\rangle \le\\ &-\sigma_h[\mathtt{L2}-\mathtt{L2min}+g(\tau_2,\tau_3)]+\\ &\epsilon[-F_{\mathrm{L2}}(2,1)+\sigma_g+F_{\mathrm{L2}}(\mathtt{MV2},\mathtt{P2})]\le\\ &-\sigma_h[\mathtt{L2max}-\mathtt{L2min}+g(\tau_2,\tau_3)+h(\tau_2,\tau_3)]+\\ &\epsilon[-F_{\mathrm{L2}}(2,1)+\sigma_g+F_{\mathrm{L2}}(\mathtt{MV2},\mathtt{P2})].\end{aligned}$$

Hence, for ϵ sufficiently small, we conclude that

$$\langle \nabla B(x_2), F_2(x_2)\rangle \le 0.$$

■

D. Modeling Actuator Attacks in SWaT

In this section, we assume that the attacker has the ability to falsify the control signals (MV1, MV2, MV3) and (P1, P2, P3) that the controllers (C1, C2, C3) send to the corresponding motor valves and pumps. That is, we let $w_i := (w_{ip}, w_{im}) \in \{0,1,2\}\times\{0,1\}$, for all $i\in\{1,2,3\}$, be the signal sent by the attacker to the i-th pump and to the i-th motor valve, respectively. In the presence of attacks, the variables (Pi,MVi) do not necessarily correspond to the actual states of the $i-$th motor valve and the $i-$th pump, respectively. For this reason, we introduce the extra variables $(\mathtt{MV1}^a,\mathtt{MV2}^a,\mathtt{MV3}^a)$ to denote the actual states of the motor valves, $(\tau_1^a,\tau_2^a,\tau_3^a)$ to time the actual transitions of the motor valves, and $(\mathtt{P1}^a,\mathtt{P2}^a,\mathtt{P3}^a)$ to denote the actual states of the pumps. Note that $\mathtt{Pi}^a = w_{ip}$ for all $i\in\{1,2,3\}$. Furthermore, to relate MVi to w_{im}, we consider the following assumption:

(A2) For each $i\in\{1,2,3\}$, if $\mathtt{MVi}^a\in\{0,1\}$, then $w_{im}\in\{\mathtt{MVi}^a,2\}$.

Under (A2), the attacker never bypasses the transition state of the motor valves. Indeed, when the attacker does not respect (A2), the device (valve) goes into a warning state and the attack would be detected immediately (this was verified in the real-world testbed). For example, if $\mathtt{MV1}^a = 1$, the attacker needs to send first the transition command $w_{1m}=2$ before sending the command $w_{1m}=0$.

Remark 3. In the absence of actuator attacks, for each $i\in\{1,2,3\}$, it follows that $w_{ip}=\mathtt{Pi}=\mathtt{Pi}^a$ and $\mathtt{MVi}=\mathtt{MVi}^a$. Furthermore, $w_{im}=\mathtt{MVi}$ if $\mathtt{MVi}\in\{0,1\}$ and $w_{im}=2$ otherwise. •

D.1. Control Logic with Adversary Model

Stage 3. The actual state of the motor valve $\mathtt{MV3}^a$ is governed by the following rules:

$$\mathtt{MV3}^a := \begin{cases}0 & \text{if } \mathtt{MV3}^a = 3 \text{ and } \tau_3^a \ge T_3\\ 1 & \text{if } \mathtt{MV3}^a = 2 \text{ and } \tau_3^a \ge T_3\\ 2 & \text{if } \mathtt{MV3}^a = 0 \text{ and } w_{3m} = 2\\ 3 & \text{if } \mathtt{MV3}^a = 1 \text{ and } w_{3m} = 2.\end{cases}$$

Moreover, we reset the value of τ_3^a to 0 whenever we switch the value of $\mathtt{MV3}^a$. In compact form, the discrete behavior of $\mathtt{MV3}^a$ and τ_3^a can be modeled by the following constrained difference equation:

$$\begin{pmatrix}\tau_3^{a+}\\ \mathtt{MV3}^{a+}\end{pmatrix} = \begin{pmatrix}0\\ G_{\mathtt{MV3}^a}(\mathtt{MV3}^a)\end{pmatrix} \qquad (\tau_3^a,\mathtt{MV3}^a,w_{3m})\in D_{\mathtt{MV3}^a},$$

$$\begin{aligned}D_{\mathtt{MV3}^a} :=&\{(\tau_3^a,\mathtt{MV3}^a,w_{3m}) : (\mathtt{MV3}^a,w_{3m})\in\{0,1\}\times\{2\}\}\cup\\ &\{(\tau_3^a,\mathtt{MV3}^a,w_{3m}) : (\mathtt{MV3}^a,\tau_3^a)\in\{2,3\}\times\{T_3\}\},\end{aligned}$$

$$G_{\mathtt{MV3}^a}(\mathtt{MV3}^a) := \begin{cases}3-\mathtt{MV3}^a & \text{if } \mathtt{MV3}^a = \{2,3\}\\ 2+\mathtt{MV3}^a & \text{if } \mathtt{MV3}^a \in\{0,1\}.\end{cases}$$

Stage 2. Similarly, the discrete behavior of $\mathtt{MV2}^a$ and τ_2^a can be modeled by the following constrained difference equation:

$$\begin{pmatrix}\tau_2^{a+}\\ \mathtt{MV2}^{a+}\end{pmatrix} = \begin{pmatrix}0\\ G_{\mathtt{MV2}^a}(\mathtt{MV2}^a)\end{pmatrix} \qquad (\tau_2^a,\mathtt{MV2}^a,w_{2m})\in D_{\mathtt{MV2}^a},$$

$$\begin{aligned}D_{\mathtt{MV2}^a} :=&\{(\tau_2^a,\mathtt{MV2}^a,w_{2m}) : (\mathtt{MV2}^a,w_{2m})\in\{0,1\}\times\{2\}\}\cup\\ &\{(\tau_2^a,\mathtt{MV2}^a,w_{2m}) : (\mathtt{MV2}^a,\tau_2^a)\in\{2,3\}\times\{T_2\}\},\end{aligned}$$

$$G_{\mathtt{MV2}^a}(\mathtt{MV2}^a) := \begin{cases}3-\mathtt{MV2}^a & \text{if } \mathtt{MV2}^a\{2,3\}\\ 2+\mathtt{MV2}^a & \text{if } \mathtt{MV2}^a \in\{0,1\}.\end{cases}$$

Stage 1. Similarly to Stage 2, the discrete behavior of $\mathtt{MV1}^a$ and τ_1^a can be modeled by the following constrained difference equation:

$$\begin{pmatrix}\tau_1^{a+}\\ \mathtt{MV1}^{a+}\end{pmatrix} = \begin{pmatrix}0\\ G_{\mathtt{MV1}^a}(\mathtt{MV1}^a)\end{pmatrix} \qquad (\tau_1^a,\mathtt{MV1}^a,w_{2m})\in D_{\mathtt{MV1}^a},$$

$$
\begin{aligned}
D_{\mathtt{MV1}^a} := \{(\tau_1^a, \mathtt{MV1}^a, w_{2m}) : (\mathtt{MV1}^a, w_{2m}) \in \{0,1\} \times \{2\}\} \cup \\
\{(\tau_1^a, \mathtt{MV1}^a, w_{2m}) : (\mathtt{MV1}^a, \tau_1^a) \in \{2,3\} \times \{T_1\}\},
\end{aligned}
$$

$$
G_{\mathtt{MV1}^a}(\mathtt{MV1}^a) := \begin{cases} 3 - \mathtt{MV1}^a & \text{if } \mathtt{MV1}^a \{2,3\} \\ 2 + \mathtt{MV1}^a & \text{if } \mathtt{MV1}^a \in \{0,1\}. \end{cases}
$$

D.2. Physics of the System Under Attack

Since the attacker can arbitrarily modify the actual state of any pump or motor valve, it follows that (25) is not guaranteed and the plant loses its cascaded interconnection. Indeed, when for example $\mathtt{MV2} = \mathtt{ON}$, the stream of the water flowing from Stage 1 to Stage 2 cannot be the same if $\mathtt{P1} = \mathtt{ON}$ or if $\mathtt{P1} = \mathtt{OFF}$. Hence, the dynamics of the water levels F_{L3} and F_{L2} in Stages 3 and 2, respectively, will additionally depend on the actual state of pump in the previous stage and the actual state of the motor valve in the next stage. In such a general scenario, the dynamics of the water levels in Stages 3 and 2 can be expressed as follows:

$$
\dot{\mathtt{L}}3 = F_{\mathrm{L3}}(\mathtt{MV3}^a, \mathtt{P2}^a), \tag{65}
$$

$$
\dot{\mathtt{L}}2 = F_{\mathrm{L2}}(\mathtt{MV2}^a, \mathtt{MV3}^a, \mathtt{P2}^a, \mathtt{P1}^a), \tag{66}
$$

where the functions F_{L3} and F_{L2} satisfy the following properties:

$$
\begin{aligned}
&F_{\mathrm{L3}}(1,1) > 0, \; F_{\mathrm{L3}}(2,1) = F_{\mathrm{L3}}(3,1) > 0, \\
&F_{\mathrm{L3}}(2,0) = F_{\mathrm{L3}}(3,0) = F_{\mathrm{L3}}(0,0) = F_{\mathrm{L3}}(0,1) < 0
\end{aligned} \tag{67}
$$

$$
\begin{aligned}
&F_{\mathrm{L2}}(\mathtt{MV2}^a, 0, 0, \mathtt{P1}^a) = F_{\mathrm{L2}}(\mathtt{MV2}^a, 0, 1, \mathtt{P1}^a) \\
&= F_{\mathrm{L2}}(\mathtt{MV2}^a, 2, 0, \mathtt{P1}^a) = F_{\mathrm{L2}}(\mathtt{MV2}^a, 1, 0, \mathtt{P1}^a) \geq 0 \\
&\qquad \forall (\mathtt{MV2}^a, \mathtt{P1}^a) \in \{0,1,2\} \times \{0,1\}, \\
&F_{\mathrm{L2}}(0, \mathtt{MV3}^a, \mathtt{P2}^a, 0) = F_{\mathrm{L2}}(0, \mathtt{MV3}^a, \mathtt{P2}^a, 1) \\
&= F_{\mathrm{L2}}(2, \mathtt{MV3}^a, \mathtt{P2}^a, 0) = F_{\mathrm{L2}}(1, \mathtt{MV3}^a, \mathtt{P2}^a, 0) \leq 0 \\
&\qquad \forall (\mathtt{MV3}^a, \mathtt{P2}^a) \in \{0,1,2\} \times \{0,1\}, \\
&F_{\mathrm{L2}}(\mathtt{MV2}^a, \mathtt{MV3}^a, 0, 0) = 0 \quad \forall (\mathtt{MV2}^a, \mathtt{MV3}^a) \in \{0,1,2\}^2, \\
&F_{\mathrm{L2}}(0, 0, \mathtt{P2}^a, \mathtt{P1}^a) = 0 \quad \forall (\mathtt{P2}^a, \mathtt{P1}^a) \in \{0,1\}^2, \\
&F_{\mathrm{L2}}(2,0,1,1) \geq F_{\mathrm{L2}}(2,2,1,1) > 0, \quad F_{\mathrm{L2}}(1,1,1,1) > 0, \\
&F_{\mathrm{L2}}(1,2,1,1) \geq 0, \quad F_{\mathrm{L2}}(2,1,1,1) \leq 0.
\end{aligned} \tag{68}
$$

In (68), we are assuming that

$$
F_{\mathrm{L2}}(\mathtt{MV2}^a, 2, \mathtt{P2}^a, \mathtt{P1}^a) = F_{\mathrm{L2}}(\mathtt{MV2}^a, 3, \mathtt{P2}^a, \mathtt{P1}^a)
$$

for all $(\mathtt{MV2}^a, \mathtt{P2}^a, \mathtt{P1}^a) \in \{0,1,2,3\} \times \{0,1\} \times \{0,1\}$ and

$$
F_{\mathrm{L2}}(2, \mathtt{MV3}^a, \mathtt{P2}^a, \mathtt{P1}^a) = F_{\mathrm{L2}}(3, \mathtt{MV3}^a, \mathtt{P2}^a, \mathtt{P1}^a)
$$

for all $(\mathtt{MV3}^a, \mathtt{P2}^a, \mathtt{P1}^a) \in \{0,1,2,3\} \times \{0,1\} \times \{0,1\}$.

Conditions (67) and (68) are important for our security results in Section 6 to hold. From the numerical values of $(F_{\mathrm{L3}}, F_{\mathrm{L2}}, F_{\mathrm{L1}})$ in Section G, we can see that these conditions are satisfied for the considered industrial plant.

D.3. Combining Logic and Physical Models under Attack

Stage 3. The new augmented state vector for Stage 3 is given by

$$
x_3^a := (x_3, \tau_3^a, \mathtt{MV3}^a) \in X_3^a := X_3 \times [0, T_3] \times \{0,1,2,3\}.
$$

Furthermore, the signal sent by the attacker to the third motor valve is $w_{3m} \in \mathcal{W}_3 := \{0,1,2\}$. Finally, the variable of the second stage affecting the dynamics of the third stage is $u_3 := \mathtt{P2}^a = w_{2p} \in \{0,1\} := \mathcal{U}_3$. The dynamical model of Stage 3 under actuator attacks is given by:

$$
\mathcal{H}_{3w} : \begin{cases} \dot{x}_3^a = F_{3w}(x_3^a, w_{3m}, u_3) \; (x_3^a, w_{3m}, u_3) \in C_{3w} \\ x_3^{a+} = G_{3w}(x_3^a, w_{3m}, u_3) \; (x_3^a, w_{3m}, u_3) \in D_{3w}, \end{cases} \tag{69}
$$

where $C_{3w} := \mathrm{cl}((X_3^a \times \mathcal{W}_3 \times \mathcal{U}_3) \backslash D_{3w})$,

$$
\begin{aligned}
D_{3w} := \{(x_3^a, w_{3m}, u_3) : x_3 \in D_3\} \cup \\
\{(x_3^a, w_{3m}, u_3) : (\tau_3^a, \mathtt{MV3}^a, w_{3m}) \in D_{\mathtt{MV3}^a}\}.
\end{aligned}
$$

Furthermore, the jump map G_{3w} is given by

$$
G_{3w}(x_3^a, w_{3m}, u_3) := \begin{bmatrix} G_{x_3}(x_3^a, w_{3m}, u_3) \\ G_{\tau_3^a}(x_3^a, w_{3m}, u_3) \\ G'_{\mathtt{MV3}^a}(x_3^a, w_{3m}, u_3) \end{bmatrix},
$$

$$
G_{x_3}(x_3^a, w_{3m}, u_3) := \begin{cases} G_3(x_3) & \text{if } x_3 \in D_3 \\ x_3 & \text{otherwise,} \end{cases}
$$

$$
G'_{\mathtt{MV3}^a}(x_3^a, w_{3m}, u_3) := \begin{cases} G_{\mathtt{MV3}^a}(\mathtt{MV3}^a) \\ \qquad \text{if } (\tau_3^a, \mathtt{MV3}^a, w_{3m}) \in D_{\mathtt{MV3}^a} \\ \mathtt{MV3}^a \qquad \text{otherwise,} \end{cases}
$$

$$
G_{\tau_3^a}(x_3^a, w_{3m}, u_3) := \begin{cases} 0 & \text{if } (\tau_3^a, \mathtt{MV3}^a, w_{3m}) \in D_{\mathtt{MV3}^a} \\ \tau_3^a & \text{otherwise.} \end{cases}
$$

Finally, the flow map F_{3w} is given by

$$
F_{3w}(x_3^a, w_{3m}, u_3) := [F_{\mathrm{L3}}(\mathtt{MV3}^a, w_{2p}) \quad 1 \quad 0 \quad 1 \quad 0]^\top.
$$

Stage 2. The new augmented state vector is given by

$$
x_2^a := (x_2, \tau_2^a, \mathtt{MV2}^a) \in X_2^a := X_2 \times [0, T_2] \times \{0,1,2,3\}.
$$

Furthermore, the signals sent by the attacker to the motor valve and the pump are $w_2 := (w_{2m}, w_{2p}) \in \mathcal{W}_2 := \{0,1,2\} \times \{0,1\}$. Finally, the variables of the first and the third stages affecting the dynamics of the second stage are $u_2 := (\mathtt{L3}, \mathtt{P1}^a, \mathtt{MV3}^a, \mathtt{MV3}) \in \mathcal{U}_2$, $\mathcal{U}_2 := \mathbb{R}_{\geq 0} \times \{0,1\} \times \{0,1,2,3\}^2$. The dynamical model under actuator attacks is given by:

$$
\mathcal{H}_{2w} : \begin{cases} \dot{x}_2^a = F_{2w}(x_2^a, w_2, u_2) & (x_2^a, w_2, u_2) \in C_{2w} \\ x_2^{a+} = G_{2w}(x_2^a, w_2, u_2) & (x_2^a, w_2, u_2) \in D_{2w}, \end{cases} \tag{70}
$$

where $C_{2w} := \mathrm{cl}((X_2^a \times \mathcal{W}_2 \times \mathcal{U}_2) \backslash D_{2w})$,

$$
\begin{aligned}
D_{2w} := \{(x_2^a, w_2, u_2) : (x_2, \mathtt{L3}, \mathtt{MV3}) \in D_2\} \cup \\
\{(x_2^a, w_{2m}, u_2) : (\tau_2^a, \mathtt{MV2}^a, w_{2m}) \in D_{\mathtt{MV2}^a}\},
\end{aligned}
$$

$$
G_{2w}(x_2^a, w_2, u_2) := \begin{bmatrix} G_{x_2}(x_2^a, w_2, u_2) \\ G_{\tau_2^a}(x_2^a, w_2, u_2) \\ G'_{\mathtt{MV2}^a}(x_2^a, w_2, u_2) \end{bmatrix},
$$

where $G_{x_2}(x_2^a, w_2, u_2) := \begin{cases} G_2(x_2, \mathtt{L3}, \mathtt{MV3}) & \text{if } (x_2, \mathtt{L3}, \mathtt{MV3}) \in D_2 \\ x_2 & \text{otherwise,} \end{cases}$

$$G'_{\mathtt{MV2}^a}(x_2^a, w_2, u_2) := \begin{cases} G_{\mathtt{MV2}^a}(\mathtt{MV2}^a) & \text{if } (\tau_2^a, \mathtt{MV2}^a, w_{2m}) \in D_{\mathtt{MV2}^a} \\ \mathtt{MV2}^a & \text{otherwise,} \end{cases}$$

$$G_{\tau_2^a}(x_2^a, w_2, u_2) := \begin{cases} 0 & \text{if } (\tau_2^a, \mathtt{MV2}^a, w_{2m}) \in D_{\mathtt{MV2}^a} \\ \tau_2^a & \text{otherwise,} \end{cases}$$

$$F_{2w}(x_2^a, w_2, u_2) := \begin{bmatrix} F_{\mathtt{L2}}(\mathtt{MV2}^a, \mathtt{MV3}^a, w_{2p}, w_{1p}) \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

D.4. Modeling Stage 3 When Only $\mathtt{P2}^a$ is Attacked

In this case, we do not need to extend x_3 nor to consider the signal w_{3m}. We only consider

$$u_3 := \mathtt{P2}^a \in \mathcal{U}_3 := \{0, 1\}$$

affecting the dynamics of Stage 3. Hence, the dynamical model of Stage 3 is given by:

$$\mathcal{H}_3 : \begin{cases} \dot{x}_3 = F_3(x_3, u_3) & (x_3, u_3) \in C_3 \times \mathcal{U}_3 \\ x_3^+ = G_3(x_3) & (x_3, u_3) \in D_3 \times \mathcal{U}_3, \end{cases} \tag{71}$$

where the flow map F_3 is given by

$$F_3(x_3, u) := [F_{\mathtt{L3}}(\mathtt{MV3}, \mathtt{P2}^a) \;\; 1 \;\; 0]^\top,$$

and $F_{\mathtt{L3}}$ satisfies (67).

D.5. Modeling Stage 3 in the Absence of Attacks and When $\mathtt{P3}$ is Hardened

In this case, the dynamics of $\mathtt{L3}$ can be expressed as follows:

$$\dot{\mathtt{L3}} = F_{\mathtt{L3}}(\mathtt{MV3}, \mathtt{P3}).$$

Furthermore, the signal $\mathtt{P3}$ that the controller $\mathtt{C3}$ sends to the third pump is governed by the logic (88). Hence, the local state vector of the third stage is $\bar{x}_3 := [x_3 \;\; \mathtt{P3}]^\top \in \bar{X}_3$, where

$$\bar{X}_3 := X_3 \times \{0, 1\}.$$

The hybrid equation modeling Stage 3 is given by:

$$\mathcal{H}_3 : \begin{cases} \dot{\bar{x}}_3 = \bar{F}_3(\bar{x}_3) & \bar{x}_3 \in \bar{C}_3 \\ \bar{x}_3^+ = \bar{G}_3(\bar{x}_3) & \bar{x}_3 \in \bar{D}_3, \end{cases} \tag{72}$$

where

$$\begin{aligned} \bar{C}_3 &:= \operatorname{cl}(\bar{X}_3 \backslash \bar{D}_3), \\ \bar{D}_3 &:= (D_3 \times \{0, 1\}) \cup \{\bar{x}_3 : (\mathtt{L3}, \mathtt{P3}) \in D_{\mathtt{P3}}\}. \end{aligned}$$

Furthermore, the jump map $\bar{G}_3$ is given by:

$$\bar{G}_3(\bar{x}_3) := [G_{x_3}(\bar{x}_3) \quad G'_{\mathtt{P3}}(\bar{x}_3)]^\top,$$

where

$$G_{x_3}(\bar{x}_3) := \begin{cases} G_3(x_3) & \text{if } x_3 \in D_3 \\ x_3 & \text{otherwise,} \end{cases}$$

$$G'_{\mathtt{P3}}(\bar{x}_3) := \begin{cases} 1 - \mathtt{P3} & \text{if } (\mathtt{L3}, \mathtt{P3}) \in D_{\mathtt{P3}} \\ \mathtt{P3} & \text{otherwise,} \end{cases}$$

and

$$\bar{F}_3(\bar{x}_3) := [F_{\mathtt{L3}}(\mathtt{MV3}, \mathtt{P3}) \;\; 1 \;\; 0 \;\; 0]^\top,$$

with

$$\begin{gathered} F_{\mathtt{L3}}(2, 0) = F_{\mathtt{L3}}(3, 0) > 0, \;\; F_{\mathtt{L3}}(2, 1) = F_{\mathtt{L3}}(3, 1) < 0, \\ F_{\mathtt{L3}}(1, \mathtt{P3}) > 0 \;\; \forall \mathtt{P3} \in \{0, 1\}, \;\; F_{\mathtt{L3}}(0, 1) < 0, \\ F_{\mathtt{L3}}(0, 0) = 0. \end{gathered} \tag{73}$$

D.6. Modeling Stage 3 Under Actuator Attacks When $\mathtt{P3}$ is Hardened

In the following, we extend the dynamical model in (69) when $\mathtt{P3}$ is not Always Open. In such a general scenario, the flow dynamics (65) in the third stage depends also on $\mathtt{P3}$ and can be expressed as follows:

$$\dot{\mathtt{L3}} = F_{\mathtt{L3}}(\mathtt{MV3}^a, \mathtt{P3}^a, \mathtt{P2}^a). \tag{74}$$

Furthermore, we still assume that the attacker affects only the actual state of the third motor valve. Hence, the new state vector is given by

$$\bar{x}_3^a := (\bar{x}_3, \tau_3^a, \mathtt{MV3}^a) \in \bar{X}_3^a := \bar{X}_3 \times [0, T_3] \times \{0, 1, 2, 3\}.$$

Furthermore, the signal sent by the attacker to the third motor valve is

$$w_{3m} \in \mathcal{W}_3 := \{0, 1, 2\}.$$

Finally, the variable of the second stage affecting the dynamics of the third stage is

$$u_3 := \mathtt{P2}^a \in \mathcal{U}_3 := \{0, 1\}.$$

The dynamical model of Stage 3 under actuator attacks is given by:

$$\mathcal{H}_{3w} : \begin{cases} \dot{\bar{x}}_3^a = \bar{F}_{3w}(\bar{x}_3^a, w_{3m}, u_3) & (\bar{x}_3^a, w_{3m}, u_3) \in \bar{C}_{3w} \\ \bar{x}_3^{a+} = \bar{G}_{3w}(\bar{x}_3^a, w_{3m}, u_3) & (\bar{x}_3^a, w_{3m}, u_3) \in \bar{D}_{3w}, \end{cases}$$

where the sets $\bar{C}_{3w}$ and $\bar{D}_{3w}$ are given by

$$\begin{aligned} \bar{C}_{3w} &:= \operatorname{cl}((\bar{X}_3^a \times \mathcal{W}_3 \times \mathcal{U}_3) \backslash \bar{D}_{3w}), \\ \bar{D}_{3w} &:= \{(\bar{x}_3^a, w_{3m}, u_3) : \bar{x}_3 \in \bar{D}_3\} \cup \\ &\qquad \{(\bar{x}_3^a, w_{3m}, u_3) : (\tau_3^a, \mathtt{MV3}^a, w_{3m}) \in D_{\mathtt{MV3}^a}\}, \end{aligned}$$

$$\bar{G}_{3w}(\bar{x}_3^a, w_{3m}, u_3) := \begin{bmatrix} G_{\bar{x}_3}(\bar{x}_3^a, w_{3m}, u_3) \\ G'_{\tau_3^a}(\bar{x}_3^a, w_{3m}, u_3) \\ G'_{\mathtt{MV3}^a}(\bar{x}_3^a, w_{3m}, u_3) \end{bmatrix},$$

where

$$G_{\bar{x}_3}(\bar{x}_3^a, w_{3m}, u_3) := \begin{cases} \bar{G}_3(\bar{x}_3) & \text{if } \bar{x}_3 \in \bar{D}_3 \\ \bar{x}_3 & \text{otherwise,} \end{cases}$$

$$G'_{\mathtt{MV3}^a}(\bar{x}_3^a, w_{3m}, u_3) := \begin{cases} G_{\mathtt{MV3}^a}(\mathtt{MV3}^a) & \text{if } (\tau_3^a, \mathtt{MV3}^a, w_{3m}) \in D_{\mathtt{MV3}^a} \\ \mathtt{MV3}^a & \text{otherwise,} \end{cases}$$

$$G'_{\tau_3^a}(\bar{x}_3^a, w_{3m}, u_3) := \begin{cases} 0 & \text{if } (\tau_3^a, \mathtt{MV3}^a, w_{3m}) \in D_{\mathtt{MV3}^a} \\ \tau_3^a & \text{otherwise.} \end{cases}$$

Finally, the flow map $\bar{F}_{3w}$ is given by

$$\bar{F}_{3w}(\bar{x}_3^a, w_{3m}, u) := [F_{\mathrm{L3}}(\mathtt{MV3}^a, \mathtt{P3}^a, \mathtt{P2}^a)\quad 1\quad 0\quad 0\quad 1\quad 0]^\top,$$

where

$$F_{\mathrm{L3}}(0, \mathtt{P3}^a, 0) = F_{\mathrm{L3}}(0, \mathtt{P3}^a, 1) = F_{\mathrm{L3}}(2, \mathtt{P3}^a, 0) = F_{\mathrm{L3}}(3, \mathtt{P3}^a, 0) = F_{\mathrm{L3}}(1, \mathtt{P3}^a, 0) \le 0,$$

$$F_{\mathrm{L3}}(\mathtt{MV3}^a, 0, \mathtt{P2}^a) \ge 0,$$

$$F_{\mathrm{L3}}(\mathtt{MV3}^a, 0, 0) = 0 \quad \forall \mathtt{MV3}^a \in \{0,1,2,3\},$$

$$F_{\mathrm{L3}}(0, 0, \mathtt{P2}^a) = 0 \quad \forall \mathtt{P2}^a \in \{0,1\},$$

$$F_{\mathrm{L3}}(\mathtt{MV3}^a, 1, 1) > 0 \quad \forall \mathtt{MV3}^a \in \{2,3\},$$

$$F_{\mathrm{L3}}(1,1,1) > 0.$$

D.7. Modeling Stage 3 When Only $\mathtt{P2}^a$ is Attacked and When $\mathtt{P3}$ is Hardened

In this case, we include $u_3 := \mathtt{P2}^a = w_{2m} \in \mathcal{U}_3 := \{0,1\}$ as an external signal affecting Stage 3. Hence, the dynamical model of Stage 3 when $\mathtt{P2}^a$ is attacked and $\mathtt{P3}$ is not Always Open is given by:

$$\mathcal{H}_3 : \begin{cases} \dot{\bar{x}}_3 = \bar{F}_3(\bar{x}_3, u_3) & (\bar{x}_3, u_3) \in \bar{C}_3 \times \mathcal{U}_3 \\ \bar{x}_3^+ = \bar{G}_3(\bar{x}_3) & (\bar{x}_3, u_3) \in \bar{D}_3 \times \mathcal{U}_3, \end{cases} \tag{75}$$

where

$$\bar{F}_3(\bar{x}_3, u_3) := [F_{\mathrm{L3}}(\mathtt{MV3}, \mathtt{P3}, \mathtt{P2}^a)\quad 1\quad 0\quad 0]^\top$$

with

$$\begin{aligned} &F_{\mathrm{L3}}(0, \mathtt{P3}^a, 0) = F_{\mathrm{L3}}(0, \mathtt{P3}^a, 1) = \\ &F_{\mathrm{L3}}(2, \mathtt{P3}^a, 0) = F_{\mathrm{L3}}(3, \mathtt{P3}^a, 0) \\ &\qquad = F_{\mathrm{L3}}(1, \mathtt{P3}^a, 0) \le 0 \quad \forall \mathtt{P3}^a \in \{0,1\}, \\ &F_{\mathrm{L3}}(\mathtt{MV3}^a, 0, \mathtt{P2}^a) \ge 0 \quad \forall (\mathtt{MV3}^a, \mathtt{P2}^a) \in \{0,1,2,3\} \times \{0,1\}, \\ &F_{\mathrm{L3}}(\mathtt{MV3}^a, 0, 0) = 0 \quad \forall \mathtt{MV3}^a \in \{0,1,2,3\}, \\ &F_{\mathrm{L3}}(0, 0, \mathtt{P2}^a) = 0 \quad \forall \mathtt{P2}^a \in \{0,1\}, \\ &F_{\mathrm{L3}}(\mathtt{MV3}^a, 1, 1) > 0 \quad \forall \mathtt{MV3}^a \in \{2,3\}, \\ &F_{\mathrm{L3}}(1,1,1) > 0. \end{aligned} \tag{76}$$

D.8. Modeling Stage 2 When Only $\mathtt{P2}^a$ is Attacked

In this case, we do not need to extend the state vector x_2. Furthermore, since the actuator attack affects $\mathtt{P2}^a$ only; it follows that

$$w_2 := w_{2p} \in \mathcal{W}_2 := \{0,1\}.$$

Finally, the variables from Stages 1 and 3 that affect the dynamics of Stage 2 are

$$\begin{aligned} u_2 &:= (\mathtt{L3}, \mathtt{MV3}) \in \mathcal{U}_2 \\ \mathcal{U}_2 &:= [\mathtt{L3min}, \mathtt{L3max} + \delta] \times \{0,1,2,3\}. \end{aligned}$$

The dynamical model under actuator attacks is given by

$$\mathcal{H}_{2w} : \begin{cases} \dot{x}_2 = F_{2w}(x_2, w_2, u_2) & (x_2, w_2, u_2) \in C_{2w} \\ x_2^+ = G_2(x_2, u_2) & (x_2, w_2, u_2) \in D_{2w}, \end{cases} \tag{77}$$

where the sets C_{2w} and D_{2w} are given by

$$\begin{aligned} C_{2w} &:= \mathrm{cl}((X_2 \times \mathcal{W}_2 \times \mathcal{U}_2) \backslash D_{2w}), \\ D_{2w} &:= \{(x_2, w_2, u_2) \in X_2 \times \mathcal{W}_2 \times \mathcal{U}_2 : (x_2, u_2) \in D_2\}. \end{aligned}$$

Finally, the flow map F_{2w} is given by

$$F_{2w}(x_2, w_2, u_2) := \begin{bmatrix} F_{\mathrm{L2}}(\mathtt{MV2}, \mathtt{MV3}, \mathtt{P2}^a = w_{2p}) \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

Using (68), and (25) when $i = 1$, we conclude that

$$\begin{aligned} &F_{\mathrm{L2}}(\mathtt{MV2}^a, 0, 0) = F_{\mathrm{L2}}(\mathtt{MV2}^a, 0, 1) = F_{\mathrm{L2}}(\mathtt{MV2}^a, 2, 0) \\ &= F_{\mathrm{L2}}(\mathtt{MV2}^a, 1, 0) \ge 0 \quad \forall \mathtt{MV2}^a \in \{0,1,2\}, \\ &F_{\mathrm{L2}}(0, 0, \mathtt{P2}^a) = 0 \quad \forall \mathtt{P2}^a \in \{0,1\}, \\ &F_{\mathrm{L2}}(2,0,1) \ge F_{\mathrm{L2}}(2,2,1) > 0, \quad F_{\mathrm{L2}}(1,1,1) > 0, \\ &F_{\mathrm{L2}}(1,2,1) \ge 0, \quad F_{\mathrm{L2}}(2,1,1) \le 0. \end{aligned} \tag{78}$$

Remark 4. Note that, in (78), we are assuming that

$$F_{\mathrm{L2}}(\mathtt{MV2}^a, 2, \mathtt{P2}^a) = F_{\mathrm{L2}}(\mathtt{MV2}^a, 3, \mathtt{P2}^a)$$

for all $(\mathtt{MV2}^a, \mathtt{P2}^a) \in \{0,1,2,3\} \times \{0,1\}$ and

$$F_{\mathrm{L2}}(2, \mathtt{MV3}^a, \mathtt{P2}^a) = F_{\mathrm{L2}}(3, \mathtt{MV3}^a, \mathtt{P2}^a)$$

for all $(\mathtt{MV3}^a, \mathtt{P2}^a) \in \{0,1,2,3\} \times \{0,1\}$. •

Remark 5. The systems in (71) and (77) are particular cases of (69) and (70), respectively, when the motor valves are not attacked; namely, when $\mathtt{MVi} = \mathtt{MVi}^a$ and $\tau_i = \tau_i^a$ for all $i \in \{2,3\}$, and when (25) is satisfied for the first stage; namely, for $i = 1$. •

D.9. Modeling Stage 2 When Only $\mathtt{P1}^a$ is Attacked

In this case, we don't need to extend the state vector x_2. Furthermore, the variables of the first and the third stages affecting the dynamics of the second stage are

$$\begin{aligned} u_2 &:= (\mathtt{L3}, \mathtt{MV3}, \mathtt{P1}^a) \in \mathcal{U}_2 \\ \mathcal{U}_2 &:= [\mathtt{L3min}, \mathtt{L3max} + \delta] \times \{0,1,2,3\} \times \{0,1\}. \end{aligned}$$

The dynamical model is given by:

$$\mathcal{H}_2 : \begin{cases} \dot{x}_2 = F_2(x_2, u_2) & (x_2, u_2) \in C_2 \times \{0,1\} \\ x_2^+ = G_2(x_2, \mathtt{L3}, \mathtt{MV3}) & (x_2, u_2) \in D_2 \times \{0,1\}, \end{cases} \tag{79}$$

where the flow map F_2 is given by

$$F_2(x_2, u_2) := \begin{bmatrix} F_{\mathrm{L2}}(\mathtt{MV2}, \mathtt{P2}, \mathtt{P1}^a) \\ 1 \\ 0 \\ 0 \end{bmatrix}, \tag{80}$$

with

$$\begin{aligned} &F_{\mathrm{L2}}(\mathtt{MV2}^a, 0, \mathtt{P1}^a) \ge 0 \quad \forall (\mathtt{MV2}^a, \mathtt{P1}^a) \in \{0,1,2,3\} \times \{0,1\}, \\ &F_{\mathrm{L2}}(0, \mathtt{P2}^a, 0) = F_{\mathrm{L2}}(0, \mathtt{P2}^a, 1) = F_{\mathrm{L2}}(2, \mathtt{P2}^a, 0) = \\ &F_{\mathrm{L2}}(3, \mathtt{P2}^a, 0) = \\ &F_{\mathrm{L2}}(1, \mathtt{P2}^a, 0) \le 0 \quad \forall \mathtt{P2}^a \in \{0,1\}, \\ &F_{\mathrm{L2}}(1,1,1) > 0, \\ &F_{\mathrm{L2}}(\mathtt{MV2}^a, 0, 0) = 0 \quad \forall \mathtt{MV2}^a \in \{0,1,2,3\}, \\ &F_{\mathrm{L2}}(0, 0, \mathtt{P1}^a) = 0 \quad \forall \mathtt{P1}^a \in \{0,1\}, \\ &F_{\mathrm{L2}}(\mathtt{MV2}^a, 1, 1) \le 0 \quad \forall \mathtt{MV2}^a \in \{2,3\}. \end{aligned} \tag{81}$$

D.10. Modeling Stage 2 in the Absence of Attacks When the Control Logic of `P2` is modified

According to (91) and compared to (17), the pump $\texttt{P2}$ closes when $\texttt{L2} \leq \texttt{L2min} - \delta$. Hence, we extend the model (27) as follows:

$$\mathcal{H}_2 : \begin{cases} \dot{x}_2 = F_2(x_2) & (x_2, u_2) \in \tilde{C}_2 \\ x_2^+ = \tilde{G}_2(x_2, u_2) & (x_2, u_2) \in \tilde{D}_2, \end{cases} \tag{82}$$

where

$$\begin{aligned} & u_2 := (\texttt{L3}, \texttt{MV3}) \in \mathcal{U}_2 \\ & \mathcal{U}_2 := [\texttt{L3min}, \texttt{L3max} + \delta] \times \{0,1,2,3\}, \\ & \tilde{C}_2 := \text{cl}((X_2 \times \mathcal{U}_2) \backslash \tilde{D}_2), \ \tilde{D}_2 := \tilde{D}_{21} \cup \tilde{D}_{22}, \end{aligned}$$

$$\begin{aligned} \tilde{D}_{21} := & \{(x_2,u_2) : \texttt{L2} \geq \texttt{L2max}, \ \texttt{MV2} = 1\} \cup \\ & \{(x_2,u_2) : \tau_2 \geq T_2, \ \texttt{MV2} \in \{2,3\}\} \cup \\ & \{(x_2,u_2) : \texttt{L2} \leq \texttt{L2min}, \ \texttt{MV2} = 0\}, \\ \tilde{D}_{22} := & \{(x_2,u_2) : \texttt{L3} \leq \texttt{L3min}, \ \texttt{P2} = 0, \ \texttt{L2} \geq \texttt{L2min}\} \cup \\ & \{(x_2,u_2) : \texttt{MV3} \in \{1,2\}, \ \texttt{P2} = 0, \ \texttt{L2} \geq \texttt{L2min}\} \cup \\ & \{(x_2,u_2) : \texttt{L3} \leq \texttt{L3min}, \ \texttt{P2} = 0, \ \texttt{L2} \geq \texttt{L2min}\} \cup \\ & \{(x_2,u_2) : \texttt{MV3} \in \{1,2\}, \ \texttt{P2} = 0, \ \texttt{L2} \geq \texttt{L2min}\} \cup \\ & \{(x_2,u_2) : \texttt{L3} \geq \texttt{L3max}, \ \texttt{P2} = 1\} \cup \\ & \{(x_2,u_2) : \texttt{MV3} \in \{0,3\}, \ \texttt{P2} = 1\} \cup \\ & \{(x_2,u_2) : \texttt{L2} \leq \texttt{L2min}, \ \texttt{P2} = 1\}. \end{aligned}$$

Furthermore, the jump map $\tilde{G}_2$ is given by:

$$\tilde{G}_2(x_2,u_2) := \begin{bmatrix} \texttt{L2} \\ G_{22}(x_2,u_2) \\ G_{23}(x_2,u_2) \\ \tilde{G}_{24}(x_2,u_2) \end{bmatrix},$$

where

$$G_{22}(x_2,u_2) := \begin{cases} 0 & \text{if } (x_2,u_2) \in \tilde{D}_{21} \\ \tau_2 & \text{otherwise,} \end{cases}$$

$$G_{23}(x_2,u_2) := \begin{cases} G_{\texttt{MV2}}(\texttt{MV2}) & \text{if } (x_2,u_2) \in \tilde{D}_{21} \\ \texttt{MV2} & \text{otherwise,} \end{cases}$$

and

$$\tilde{G}_{24}(x_2,u_2) := \begin{cases} 1 - \texttt{P2} & \text{if } (x_2,u_2) \in \tilde{D}_{22} \\ \texttt{P2} & \text{otherwise.} \end{cases}$$

D.11. Modeling Stage 2 When Only `P1`^{a} is Attacked and the Control Logic of `P2` is Modified

The dynamical model is given by:

$$\mathcal{H}_2 : \begin{cases} \dot{x}_2 = F_2(x_2,u_2) & (x_2,u_2) \in \tilde{C}_2 \times \{0,1\} \\ x_2^+ = \tilde{G}_2(x_2, \texttt{L3}, \texttt{MV3}) & (x_2,u_2) \in \tilde{D}_2 \times \{0,1\}, \end{cases} \tag{83}$$

where

$$\begin{aligned} & u_2 := (\texttt{L3}, \texttt{MV3}, \texttt{P1}^a) \in \mathcal{U}_2, \\ & \mathcal{U}_2 := [\texttt{L3min}, \texttt{L3max} + \delta] \times \{0,1,2,3\} \times \{0,1\}, \end{aligned}$$

and F_2 as in (80).

E. Security Proofs Under Actuation Attacks

As our first contribution, we now adapt our previous results to reason about safety under attacks. First, we consider a hybrid system under general attacks

$$\mathcal{H}_u : \begin{cases} \dot{x} = F(x,u) & (x,u) \in C \\ x^+ = G(x,u) & (x,u) \in D. \end{cases} \tag{84}$$

Where the attack u can affect the physical states, as well as the discrete software logic. To analyze safety in the presence of attacks, we introduce a new concept we call uniform safety.

Definition 8 (Uniform Safety). System $\mathcal{H}_u$ in (84) is said to be safe with respect to (X_o, X_u) uniformly in $u \in \mathcal{U}$ iff, for each solution pair (x,u) to $\mathcal{H}_u$ such that $x(0,0) \in X_o$, the solution x never reaches the set X_u. •

In the presence of attacks, the variables $(\texttt{Pi}, \texttt{MVi})$ do not necessarily correspond to the actual states of the $i-$th motor valve and the $i-$th pump, respectively. For this reason, we introduce the extra variables $(\texttt{MV1}^a, \texttt{MV2}^a, \texttt{MV3}^a)$ to denote the actual states of the motor valves, $(\tau_1^a, \tau_2^a, \tau_3^a)$ to time the actual transitions of the motor valves, and $(\texttt{P1}^a, \texttt{P2}^a, \texttt{P3}^a)$ to denote the actual states of the pumps.

As we show in the experimental results, the system with the original PLC programs is unsafe in the presence attacks. Indeed, due to the constant demand of water by Stage 3, $\texttt{L2}$ becomes less than $\texttt{L2min} - \delta$ if $\texttt{P1}^a = 0$ is maintained by the attacker. Similarly, $\texttt{L3}$ becomes less than $\texttt{L3min}$ if $\texttt{P2}^a = 0$ is maintained. Therefore the original system is unsafe to attacks that can compromise either the first or the second pump. However, as we will show in this section, if we change the control logic of PLCs, the system can be made safe against arbitrary attacks (as long as they compromise only one control signal).

Claim 3. When the attacker forces $\texttt{P1}^a = 0$ (closes the pump in Stage 1) and/or forces $\texttt{P2}^a = 0$ (closes the pump in Stage 2), the plant becomes unsafe. •

We prove Claim 3 by finding a counterexample. Indeed, due to the constant demand of water by Stage 3, $\texttt{L2}$ becomes less than $\texttt{L2min} - \delta$ if $\texttt{P1}^a = 0$ is maintained by the attacker. Similarly, $\texttt{L3}$ becomes less than $\texttt{L3min}$ if $\texttt{P2}^a = 0$ is maintained.

Claim 4. The plant remains safe if the attacker forces $\texttt{P2}^a = 1$ and/or $\texttt{P1}^a = 1$. •

To prove Claim 4, we will formally show that the plant remains safe when the attacker enforces $\texttt{P2}^a = 1$. Intuitively, in this case, the flow of water is governed by the motor valves and the behavior of the plant is not very different from its behavior in the absence of attacks. The only change concerns the dynamics of $(\texttt{L1}, \texttt{L2}, \texttt{L3})$, which as we shall show, does not compromise the safety of the plant. We model Stages 3 and 2 when $\texttt{P2}^a$ is attacked; see system $\mathcal{H}_3$ and system system $\mathcal{H}_2$ Theorems 10 and 11 prove Claim 2 by showing safety of the plant when the attacker forces $\texttt{P2}^a = 1$.

Theorem 10. Consider the hybrid system $\mathcal{H}_3$ with $u_3 = \texttt{P2}^a = 1$. Consider the initial set X_{o3} in (30) and an unsafe set $X_{u3} \subset X_3 \backslash X_{s3}$ with X_{s3} introduced in (36). Assume that there exits $\sigma > 0$ such that

$$4T_3(F_{\mathrm{L3}}(3,1) + \sigma) \le \delta. \tag{85}$$

Then, the hybrid system $\mathcal{H}_3$ with $u_3 = \texttt{P2}^a = 1$ is safe with respect to (X_{o3}, X_{u3}), and admits a barrier function certificate given by

$$B(x_3) := (\texttt{L3} - \texttt{L3min})(\texttt{L3} - \texttt{L3max} - f(\tau_3, \texttt{MV3})),$$

where $f(\tau_3, \texttt{MV3}) := (F_{\mathrm{L3}}(3,1) + \sigma)[\tau_3 + T_3 * w_f(\texttt{MV3})]$, where $w_f(2) := 0,\ w_f(1) := 1,\ w_f(3) := 2,\ w_f(0) := 3.$ □

Proof. It can be derived by following the exact same steps as in the proof of Theorem 6 while noting, under (67), that

$$F_{\mathrm{L3}}(1,1) > 0,\ F_{\mathrm{L3}}(2,1) = F_{\mathrm{L3}}(3,1) > 0,\ F_{\mathrm{L3}}(0,1) < 0.$$

■

Theorem 11. Consider the hybrid system $\mathcal{H}_2$. Consider the initial set X_{o2} in (31) and the unsafe set $X_{u2} \subset X_2 \backslash X_{s2}$ with X_{s2} introduced in (35). Assume that there exist $\sigma_h > 0$ and $\sigma_g > 0$ such that

$$4T_2(F_{\mathrm{L2}}(2,0,0) + \sigma_h) \le \delta, \tag{86}$$
$$4T_2(F_{\mathrm{L2}}(2,1,1) + \sigma_g) \le \delta. \tag{87}$$

Then, the hybrid system $\mathcal{H}_2$ is safe with respect to (X_{o2}, X_{u2}) uniformly in $(u_2, w_2) \in \mathcal{U}_2 \times \mathcal{W}_2$, and admits a barrier function certificate given by

$$B(x_2) := (\texttt{L2} - \texttt{L2min} + g(\tau_2, \texttt{MV2})) \times (\texttt{L2} - \texttt{L2max} - h(\tau_2, \texttt{MV2})),$$

where $g(\tau_2, \texttt{MV2}) := (-F_{\mathrm{L2}}(2,1,1) + \sigma_g) * [\tau_2 + T_2 * w_g(\texttt{MV2})]$,
$w_g(3) := 0,\ w_g(0) := 1,\ w_g(2) := 2,\ w_g(1) := 3,$
$h(\tau_2, \texttt{MV2}) := (F_{\mathrm{L2}}(2,0,0) + \sigma_h) * [\tau_2 + T_2 * w_h(\texttt{MV2})]$,
$w_h(2) := 0,\ w_h(1) := 1,\ w_h(3) := 2,\ w_h(0) := 3.$
□

Proof. We note that the system in (77) can be expressed as follows:

$$\mathcal{H}_{2w} : \begin{cases} \dot{x}_2 = F_{2w}(x_2, w_2, u_2) \\ \qquad (x_2, u_2, w_2) \in C_2 \times \mathcal{W}_2 \\ x_2^+ = G_2(x_2, u_2) \\ \qquad (x_2, u_2, w_2) \in D_2 \times \mathcal{W}_2. \end{cases}$$

The rest of the proof follows exactly using the same steps as in the proof of Theorem 7 while noting, under (78), that

$$F_{\mathrm{L2}}(\texttt{MV2}, \texttt{MV3}, \texttt{P2}^a) \ge F_{\mathrm{L2}}(2,1,1)$$

for all $(\texttt{MV2}, \texttt{MV3}, \texttt{P2}^a) \in \{1,2\} \times \{0,1,2\} \times \{0,1\}$, and

$$F_{\mathrm{L2}}(\texttt{MV2}, \texttt{MV3}, \texttt{P2}^a) \le F_{\mathrm{L2}}(2,0,0)$$

for all $(\texttt{MV2}, \texttt{MV3}, \texttt{P2}^a) \in \{0,2\} \times \{0,1,2\} \times \{0,1\}$. ■

E.1. Changing the Control Logic of PLCs to Make the System More Secure

In this section we harden the system to make it more resilient to attacks. In particular, we first change the control logic of the PLC controlling stage 3 (`C3`) so that `P3` is not always 1. As a result, we include `P3` as a control parameter governed by the following logic:

$$\texttt{P3} := \begin{cases} 0 & \text{if } (\texttt{L3} \le \texttt{L3o},\ \texttt{P3} = 1) \\ 1 & \text{if } (\texttt{L3} \ge \texttt{L3o},\ \texttt{P3} = 0), \end{cases} \tag{88}$$

where $\texttt{L3o} > 0$ is a lower bound on the water level `L3` in Stage 3, it aims to avoid the dry-runs (operates without liquid) of the pump `P3`. Hence, the behavior of `P3` can be modeled by the following constrained difference equation:

$$\texttt{P3}^+ = G_{\texttt{P3}}(\texttt{P3}) \qquad (\texttt{L3}, \texttt{P3}) \in D_{\texttt{P3}},$$

where $D_{\texttt{P3}} := \{(\texttt{L3}, \texttt{P3}) : \texttt{L3} \le \texttt{L3o},\ \texttt{P3} = 1\} \cup \{(\texttt{L3}, \texttt{P3}) : \texttt{L3} \ge \texttt{L3o},\ \texttt{P3} = 0\}$ and $G_{\texttt{P3}}(\texttt{P3}) := 1 - \texttt{P3}$.

Using the logic (88) with $\texttt{L3o} = \texttt{L3min}$, we are able to show the following claim.

Claim 5. When `P3` is governed by (88) with $\texttt{L3o} = \texttt{L3min}$, the plant remains safe under any arbitrary time series of possible attacks affecting $\texttt{P2}^a$. •

To show Claim 5, we use Theorem 11 to conclude that it is enough to show safety of Stage 3 uniformly in $\texttt{P2}^a \in \{0,1\}$ when `P3` is governed by (88). To simplify the analysis, we model Stage 3 when only $\texttt{P2}^a$ is attacked and `P3` is controlled by our modified control logic (it is hardened); see system $\mathcal{H}_3$.

Theorem 12. Consider the hybrid system $\mathcal{H}_3$. Consider the initial set $\bar{X}_{o3} := X_{o3} \times \{0,1\}$ and an unsafe set $X_{u3} \subset \bar{X}_3 \backslash \bar{X}_{s3}$ with $\bar{X}_{s3} = X_{s3} \times \{0,1\}$. Assume that there exits $\sigma > 0$ such that

$$4T_3(F_{\mathrm{L3}}(3,0,1) + \sigma) \le \delta. \tag{89}$$

Then, the hybrid system $\mathcal{H}_3$ is safe with respect to $(\bar{X}_{o3}, \bar{X}_{u3})$ uniformly in $u_3 = w_{2m} = \texttt{P2}^a \in \mathcal{U}_3$, and admits a barrier function certificate given by

$$B(\bar{x}_3) := (\texttt{L3} - \texttt{L3min})(\texttt{L3} - \texttt{L3max} - \texttt{P3} * f(\tau_3, \texttt{MV3})),$$

where $f(\tau_3, \texttt{MV3}) := (F_{\mathrm{L3}}(3,0,1) + \sigma)[\tau_3 + T_3 * w_f(\texttt{MV3})]$, and $w_f(2) := 0,\ w_f(1) := 1,\ w_f(3) := 2,\ w_f(0) := 3.$ □

Proof.

Note that

$$\bar{D}_3 := \bar{D}_{31} \cup (D_3 \times \{0,1\}),$$

where

$$\begin{aligned} \bar{D}_{31} := &\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \le \texttt{L3min},\ \texttt{P3} = 1\} \cup \\ &\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \ge \texttt{L3min},\ \texttt{P3} = 0\}. \end{aligned}$$

Furthermore, we introduce the set

$$\begin{aligned} \bar{C}_3 &:= \mathrm{cl}(\bar{X}_3 \backslash (\bar{D}_{31} \cup (D_3 \times \{0,1\}))), \\ &= \mathrm{cl}(\bar{X}_3 \backslash \bar{D}_{31}) \cap \mathrm{cl}(\bar{X}_3 \backslash (D_3 \times \{0,1\})) \\ &= \mathrm{cl}(\bar{X}_3 \backslash \bar{D}_{31}) \cap (\mathrm{cl}(X_3 \backslash D_3) \times \{0,1\}) \\ &= \mathrm{cl}(\bar{X}_3 \backslash \bar{D}_{31}) \cap (C_3 \times \{0,1\}). \end{aligned}$$

Finally, we let

$$\begin{aligned}\bar{C}_{31} :=& \operatorname{cl}(\bar{X}_3 \backslash \bar{D}_{31})\\ =&\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \geq \texttt{L3min},\ \texttt{P3} = 1\} \cup\\ &\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \leq \texttt{L3min},\ \texttt{P3} = 0\},\end{aligned}$$

to conclude that

$$\bar{C}_3 = \bar{C}_{31} \cap (C_3 \times \{0,1\}).$$

Furthermore, note that

$$f(\tau_3, \texttt{MV3}) \in [0,\ (F_{\texttt{L3}}(3,0,1) + \sigma) * 4 * T_3]$$

for all $(\tau_3, \texttt{MV3}) \in [0, T_3] \times \{0,1,2,3\}$. Hence, for $\sigma > 0$ and $\delta > 0$ such that (89) holds, we conclude that

$$f(\tau_3, \texttt{MV3}) \in [0, \delta] \quad \forall (\tau_3, \texttt{MV3}) \in [0, T_3] \times \{0,1,2,3\}.$$

Thus, (4) is satisfied. Next, we let the set

$$K_e := \{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \in [\texttt{L3min}, \texttt{L3max}{+}\texttt{P3}{*}f(\tau_3, \texttt{MV3})]\}.$$

To complete the proof, we use Theorem 1 and we start verifying the jump conditions (12) and (13). Note that the set $K_e \cap \bar{D}_3$ satisfies

$$K_e \cap \bar{D}_3 = A_1 \cup A_2 \cup A_3 \cup A_4,$$

where

$$\begin{aligned}A_1 :=&\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \in [\texttt{L3max}, \texttt{L3max} + \texttt{P3} * f(\tau_3, 1)],\\ &\texttt{MV3} = \{1\},\ \texttt{P3} = 1\},\\ A_2 :=&\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \in [\texttt{L3min}, \texttt{L3max} + \texttt{P3} * f(T_3, \texttt{MV3})],\\ &\tau_3 = T_3,\ \texttt{MV3} \in \{2,3\},\ \texttt{P3} = 1\},\\ A_3 :=&\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} = \texttt{L3min},\ \texttt{P3} = 1\},\\ A_4 :=&\{\bar{x}_3 \in \bar{X}_3 \backslash (A_1 \cup A_2 \cup A_3) :\\ &\texttt{L3} \in [\texttt{L3min}, \texttt{L3max} + \texttt{P3} * f(T_3, \texttt{MV3})],\ \texttt{P3} = 0\}.\end{aligned}$$

Note that, for each $\bar{x}_3 \in A_1$, we have

$$\bar{G}_3(\bar{x}_3) = [\texttt{L3} \quad 0 \quad 3 \quad \texttt{P3}]^\top.$$

Hence,

$$\begin{aligned}&B(\bar{G}_3(\bar{x}_3)) =\\ &(\texttt{L3} - \texttt{L3min})(\texttt{L3} - \texttt{L3max} - \texttt{P3} * f(0,3)) \leq 0.\end{aligned}$$

The latter inequality is true since

$$f(\tau_3, 1) \leq f(0,3) \quad \forall \tau_3 \in [0, T_3].$$

Similarly, for each $\bar{x}_3 \in A_2$, we have

$$\bar{G}_3(\bar{x}_3) = [\texttt{L3} \quad 0 \quad \alpha(\texttt{MV3}) \quad \texttt{P3}]^\top,$$

where $\alpha(3) := 0$ and $\alpha(2) := 1$. Hence,

$$\begin{aligned}&B(\bar{G}_3(\bar{x}_3)) =\\ &(\texttt{L3} - \texttt{L3min})(\texttt{L3} - \texttt{L3max} - \texttt{P3} * f(0, \alpha(\texttt{MV3}))) \leq 0.\end{aligned}$$

The latter inequality is true since

$$f(\tau_3, 3) \leq f(0,0) \quad \forall \tau_3 \in [0, T_3],$$

and

$$f(\tau_3, 2) \leq f(0,1) \quad \forall \tau_3 \in [0, T_3].$$

Next, for each $\bar{x}_3 \in A_3$, we have

$$\bar{G}_3(\bar{x}_3) = [\texttt{L3min} \quad \star \quad \star \quad 0]^\top.$$

Hence, $B(\bar{G}_3(\bar{x}_3)) = 0$. Finally, for each $\bar{x}_3 \in A_4$, we have

$$\bar{G}_3(\bar{x}_3) = [\texttt{L3} \quad \star \quad \star \quad 1]^\top,$$

$$B(\bar{x}_3) = (\texttt{L3} - \texttt{L3min})(\texttt{L3} - \texttt{L3max}) \leq 0.$$

Hence,

$$B(\bar{G}_3(\bar{x}_3)) = (\texttt{L3}{-}\texttt{L3min})(\texttt{L3}{-}\texttt{L3max}{-}f(\tau_3, \texttt{MV3})) \leq 0.$$

We conclude that (12) is satisfied. Moreover, to show (13), we notice that $\bar{C}_3 \cup \bar{D}_3 = \bar{X}_3$ and $\bar{G}_3(\bar{x}_3) \in \bar{X}_3$ for all $\bar{x}_3 \in \bar{D}_3$.

Next, to verify (11), we start noting that the set $U(\partial K_e) \setminus K_e$ satisfies

$$U(\partial K_e) \setminus K_e = B_1 \cup B_2,$$

where, for some $\epsilon > 0$,

$$B_1 := \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \in (\texttt{L3min} - \epsilon, \texttt{L3min})\right\},$$

and

$$B_2 := \{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \in (\texttt{L3max} + \texttt{P3} * f(\tau_3, \texttt{MV3}))[1, \epsilon)\}.$$

Furthermore, the set $\bar{C}_3$ can be explicitly expressed as

$$\bar{C}_3 = (C_3 \times \{0,1\}) \cap \bar{C}_{31} = \left(\bigcap_{i=1}^{3} \bar{D}_{3i}\right) \cap \bar{C}_{31},$$

where

$$\begin{aligned}\bar{D}_{31} &:= \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \geq \texttt{L3min}\ \cup\ \texttt{MV3} \in \{1,2,3\}\right\}\\ \bar{D}_{32} &:= \left\{\bar{x}_3 \in \bar{X}_3 : \tau_3 \leq T_3\ \cup\ \texttt{MV3} \in \{0,1\}\right\}\\ \bar{D}_{33} &:= \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \leq \texttt{L3max}\ \cup\ \texttt{MV3} \in \{0,2,3\}\right\},\end{aligned}$$

This is equivalent to

$$\bar{C}_3 := \left(\bigcup_{i=1}^{4} C_{3i}\right) \cap \bar{C}_{31}, \tag{90}$$

where

$$\begin{aligned}C_{31} &:= \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3min} \leq \texttt{L3} \leq \texttt{L3max}\right\}\\ C_{32} &:= \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \leq \texttt{L3max},\ \texttt{MV3} \in \{1,2,3\}\right\}\\ C_{33} &:= \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \geq \texttt{L3min},\ \texttt{MV3} \in \{0,2,3\}\right\}\\ C_{34} &:= \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{MV3} \in \{2,3\}\right\}.\end{aligned}$$

We can also show that

$$\bar{C}_3 := C_{31}^a \cup C_{31}^b \cup C_{32}^a \cup C_{32}^b \cup C_{33}^a \cup C_{33}^b \cup C_{34}^a \cup C_{34}^b,$$

where

$$\begin{aligned}C_{31}^a &:= \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3min} \leq \texttt{L3} \leq \texttt{L3max},\ \texttt{P3} = 1\right\},\\ C_{31}^b &:= \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} = \texttt{L3min},\ \texttt{P3} = 0\right\},\\ C_{32}^a &:= \{\bar{x}_3 \in \bar{X}_3 :\\ &\qquad \texttt{L3} \in [\texttt{L3min}, \texttt{L3max}],\ \texttt{MV3} \in \{1,2,3\},\ \texttt{P3} = 1\},\\ C_{32}^b &:= \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \leq \texttt{L3min},\ \texttt{MV3} \in \{1,2,3\},\ \texttt{P3} = 0\right\},\\ C_{33}^a &:= \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \geq \texttt{L3min},\ \texttt{MV3} \in \{0,2,3\},\ \texttt{P3} = 1\right\},\\ C_{33}^b &:= \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} = \texttt{L3min},\ \texttt{MV3} \in \{0,2,3\},\ \texttt{P3} = 0\right\},\\ C_{34}^a &:= \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \leq \texttt{L3min},\ \texttt{MV3} \in \{2,3\},\ \texttt{P3} = 0\right\},\\ C_{34}^b &:= \left\{\bar{x}_3 \in \bar{X}_3 : \texttt{L3} \geq \texttt{L3min},\ \texttt{MV3} \in \{2,3\},\ \texttt{P3} = 1\right\}.\end{aligned}$$

Next, we note that

$$(U(\partial K_e)\setminus K_e)\cap \bar{C}_3 = \\ (B_1\cap C^b_{32})\cup(B_2\cap C^a_{33})\cup(B_1\cap C^a_{34})\cup(B_2\cap C^b_{34}),$$

with

$$B_1\cap C^b_{32} = \{\bar{x}_3\in\bar{X}_3 : \\ \mathtt{L3}\in(\mathtt{L3min}-\epsilon,\mathtt{L3min}),\ \mathtt{MV3}\in\{1,2,3\},\ \mathtt{P3}=0\},$$

$$B_2\cap C^a_{33} = \{\bar{x}_3\in\bar{X}_3 : \\ \mathtt{L3}\in(\mathtt{L3max}+f(\tau_3,\mathtt{MV3}),\mathtt{L3max}+f(\tau_3,\mathtt{MV3})+\epsilon), \\ \mathtt{MV3}\in\{0,2,3\},\ \mathtt{P3}=1\},$$

$$B_1\cap C^a_{34} = \{\bar{x}_3\in\bar{X}_3 : \\ \mathtt{L3}\in(\mathtt{L3min}-\epsilon,\mathtt{L3min}),\ \mathtt{MV3}\in\{2,3\},\ \mathtt{P3}=0\},$$

$$B_2\cap C^b_{34} = \{\bar{x}_3\in\bar{X}_3 : \\ \mathtt{L3}\in(\mathtt{L3max}+f(\tau_3,\mathtt{MV3}),\mathtt{L3max}+f(\tau_3,\mathtt{MV3})+\epsilon), \\ \mathtt{MV3}\in\{2,3\},\ \mathtt{P3}=1\}.$$

Next, we evaluate the product $\langle\nabla B(\bar{x}_3),\bar{F}_3(\bar{x}_3,u_3)\rangle$ at each $(\bar{x}_3,u_3)\in((U(\partial K_e)\setminus K_e)\cap\bar{C}_3)\times\{0,1\}$. Note that

$$\nabla B(\bar{x}_3) = \\ \begin{bmatrix} 2\mathtt{L3}-(\mathtt{L3max}+\mathtt{L3min})-f(\tau_3,\mathtt{MV3}) \\ -\mathtt{P3}*(F_{\mathtt{L3}}(3,0,1)+\sigma)(\mathtt{L3}-\mathtt{L3min}) \\ \star \\ \star \end{bmatrix}.$$

Hence,

$$\langle\nabla B(\bar{x}_3),\bar{F}(\bar{x}_3,u_3)\rangle = \\ F_{\mathtt{L3}}(\mathtt{MV3},\mathtt{P3},\mathtt{P2}^a)(\mathtt{L3}-\mathtt{L3max}-f(\tau_3,\mathtt{MV3}))+ \\ (\mathtt{L3}-\mathtt{L3min})(F_{\mathtt{L3}}(\mathtt{MV3},\mathtt{P3},\mathtt{P2}^a)-\mathtt{P3}*[F_{\mathtt{L3}}(3,0,1)+\sigma]).$$

Next, we distinguish the following two situations:

1) When $\bar{x}_3\in(B_1\cap C^b_{32})\cup(B_1\cap C^a_{34})$, we conclude that $\mathtt{P3}=0$, $|\mathtt{L3}-\mathtt{L3min}|\le\epsilon$, $\mathtt{L3max}-\mathtt{L3}>\mathtt{L3max}-\mathtt{L3min}$, and

$$F_{\mathtt{L3}}(\mathtt{MV3},0,\mathtt{P2}^a)\ge 0 \quad \forall(\mathtt{MV3},\mathtt{P2}^a)\in\{0,1,2\}\times\{0,1\}.$$

Hence,

$$\langle\nabla B(\bar{x}_3),\bar{F}_3(\bar{x}_3,u_3)\rangle \le F_{\mathtt{L3}}(\mathtt{MV3},0,\mathtt{P2}^a)(\mathtt{L3}-\mathtt{L3max}) \\ +\epsilon|F_{\mathtt{L3}}(\mathtt{MV3},0,\mathtt{P2}^a)| \\ \le -|F_{\mathtt{L3}}(\bar{x}_3,0,\mathtt{P2}^a)|*(\mathtt{L3max}-\mathtt{L3min}-\epsilon).$$

Hence, for ϵ sufficiently small, we conclude that, for each $(\bar{x}_3,u_3)\in[(B_1\cap C^b_{32})\cup(B_1\cap C^a_{34})]\times\{0,1\}$,

$$\langle\nabla B(\bar{x}_3),\bar{F}_3(\bar{x}_3,u_3)\rangle\le 0.$$

2) When $\bar{x}_3\in(B_2\cap C^a_{33})\cup(B_2\cap C^b_{34})$, we conclude that $\mathtt{P3}=1$, $|\mathtt{L3}-\mathtt{L3max}-f(\tau_3,\mathtt{MV3})|\le\epsilon$, $\mathtt{L3}-\mathtt{L3min}\ge\mathtt{L3max}-\mathtt{L3min}$, and

$$F_{\mathtt{L3}}(\mathtt{MV3},1,\mathtt{P2}^a)-F_{\mathtt{L3}}(3,0,1)-\sigma\le-\sigma$$

for all $(\mathtt{MV3},\mathtt{P2}^a)\in\{0,2,3\}\times\{0,1\}$. Hence,

$$\langle\nabla B(\bar{x}_3),\bar{F}_3(\bar{x}_3,u_3)\rangle\le|F_{\mathtt{L3}}(\mathtt{MV3},1,\mathtt{P2}^a)|\epsilon+ \\ (\mathtt{L3}-\mathtt{L3min})[F_{\mathtt{L3}}(\mathtt{MV3},1,\mathtt{P2}^a)-(F_{\mathtt{L3}}(3,0,1)+\sigma)] \\ \le|F_{\mathtt{L3}}(\mathtt{MV3},1,\mathtt{P2}^a)|*\epsilon-\sigma*(\mathtt{L3max}-\mathtt{L3min}).$$

Hence, for ϵ sufficiently small, we conclude that, for each $(\bar{x}_3,u_3)\in[(B_2\cap C_{33})\cup(B_2\cap C_{34})]\times\{0,1\}$,

$$\langle\nabla B(\bar{x}_3),\bar{F}_3(\bar{x}_3,u_3)\rangle\le 0.$$

■

So far, we showed that our control logic modification makes the system safe against attacks in P2, but in experimental results our change is not enough when the adversary attacks P1. As a result, we need to modify the PLC controlling stage 2 as well; i.e., C2.

E.2. Changing the Control Logic of P3 and P2 to Make the System More Secure

Claim 6. When modifying the logic in (17) governing P2 as follows:

$$\mathtt{P2} := \begin{cases} 1 \text{ if} & (\mathtt{L3}\le\mathtt{L3min},\mathtt{P2}=0,\mathtt{L2}\ge\mathtt{L2min})\vee \\ & (\mathtt{MV3}\in\{1,2\},\mathtt{P2}=0,\mathtt{L2}\ge\mathtt{L2min}), \\ 0 \text{ if} & (\mathtt{L3}\ge\mathtt{L3max},\mathtt{P2}=1)\vee \\ & (\mathtt{MV3}\in\{0,3\},\mathtt{P2}=1)\vee \\ & (\mathtt{L2}\le\mathtt{L2min},\mathtt{P2}=1), \end{cases} \tag{91}$$

the plant becomes safe under any attack affecting $\mathtt{P1}^a$. •

Since only $\mathtt{P1}^a$ is attacked, the model of Stage 3 is as in (26) and its safety is already analyzed in Theorem 6. Hence, to prove Claim 6, it is enough to prove that Stage 2 is safe uniformly in $(\mathtt{P1}^a,x_3)\in\{0,1\}\times X_{s3}$ when only $\mathtt{P1}^a$ is attacked and when (91) governs P2. To simplify the proof, we model Stage 2 when only $\mathtt{P1}^a$ is attacked and when the logic governing P2 is modified;

$$\mathcal{H}_2 : \begin{cases} \dot{x}_2 = F_2(x_2,u_2) & (x_2,u_2)\in\tilde{C}_2\times\{0,1\} \\ x_2^+ = \tilde{G}_2(x_2,\mathtt{L3},\mathtt{MV3}) & (x_2,u_2)\in\tilde{D}_2\times\{0,1\}, \end{cases} \tag{92}$$

where

$$u_2 := (\mathtt{L3},\mathtt{MV3},\mathtt{P1}^a)\in\mathcal{U}_2, \\ \mathcal{U}_2 := [\mathtt{L3min},\mathtt{L3max}+\delta]\times\{0,1,2,3\}\times\{0,1\},$$

and

$$F_2(x_2,u_2) := \begin{bmatrix} F_{\mathtt{L2}}(\mathtt{MV2},\mathtt{P2},\mathtt{P1}^a) \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

Theorem 13. Consider the hybrid system $\mathcal{H}_2$ in (92). Consider the initial set X_{o2} in (31) and the unsafe set $X_{u2}\subset X_2\backslash X_{s2}$ with X_{s2} introduced in (35). Assume that there exists $\sigma_h>0$ such that

$$4T_2(F_{\mathtt{L2}}(2,0,1)+\sigma_h)\le\delta. \tag{93}$$

Then, the hybrid system $\mathcal{H}_2$ in (92) is safe with respect to (X_{o2},X_{u2}) uniformly in $u_2\in\mathcal{U}_2$, and admits a barrier function certificate given by

$$B(x_2) := (\mathtt{L2}-\mathtt{L2min})(\mathtt{L2}-\mathtt{L2max}-\chi(\mathtt{L2})*h(\tau_2,\mathtt{MV2})), \tag{94}$$

where $\chi : \mathbb{R} \to [0,1]$ is a smooth function such that

$$\begin{cases} \chi(\texttt{L2}) = 1 & \text{if } \texttt{L2} \geq \texttt{L2max} \\ \chi(\texttt{L2}) = 0 & \text{if } \texttt{L2} \leq \texttt{L2min} \\ \chi(\texttt{L2}) \in [0,1] & \quad \text{otherwise,} \end{cases}$$

and $h(\tau_2, \texttt{MV2}) := (F_{\texttt{L2}}(2,0,1) + \sigma_h) * [\tau_2 + T_2 * w_h(\texttt{MV2})]$, $w_h(2) := 0$, $w_h(1) := 1$, $w_h(3) := 2$, $w_h(0) := 3$. □

Proof. Consider the hybrid system $\mathcal{H}_2$ in (83) with $u_2 \in \mathcal{U}_2$. Consider the sets (X_{o2}, X_{u2}) and the barrier function candidate in (94). Note that

$$h(\tau_2, \texttt{MV2}) \in [0,\ ([F_{\texttt{L2}}(2,0,1) + \sigma_h] * T_2 * 4)]$$

for all $(\tau_2, \texttt{MV2}) \in [0, T_2] \times \{0,1,2,3\}$. Hence, for σ_h satisfying (87), we conclude that

$$h(\tau_2, \texttt{MV2}) \in [0, \delta] \quad \forall (\tau_2, \texttt{MV2}) \in [0, T_2] \times \{0,1,2,3\}.$$

Thus, (4) is satisfied.

Next, we introduce the notation

$$\begin{aligned} \bar{u}_2 &:= (\texttt{L3}, \texttt{MV3}) \in \bar{\mathcal{U}}_2, \\ \bar{\mathcal{U}}_2 &:= [\texttt{L3min}, \texttt{L3max} + \delta] \times \{0,1,2,3\}. \end{aligned}$$

We also introduce

$$K_e := \{x_2 \in X_2 : \texttt{L2} \in [\texttt{L2min}, \texttt{L2max} + h(\tau_2, \texttt{MV2})]\} \times \bar{\mathcal{U}}_2.$$

Note that

$$K_w = K_e \times \{0,1\}.$$

To apply Theorem 1, we start verifying (12) and (13).

Note that the set $K_e \cap \tilde{D}_2$ satisfies

$$K_e \cap \tilde{D}_2 = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6,$$

where

$$\begin{aligned} A_1 := \big\{ &(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \\ &\texttt{L2} \in [\texttt{L2max}, \texttt{L2max} + h(\tau_2, 1)],\ \texttt{MV2} = 1 \big\}, \end{aligned}$$

$$A_2 := \big\{ (x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \texttt{L2} = \texttt{L2min},\ \texttt{MV2} = 0 \big\},$$

$$\begin{aligned} A_3 :=\{ &(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \\ &\texttt{L2} \in [\texttt{L2min}, \texttt{L2max} + h(T_2, \texttt{MV2})], \\ &\tau_2 = T_2,\ \texttt{MV2} \in \{2,3\}\}, \end{aligned}$$

$$\begin{aligned} A_4 :=\{ &(x_2, \bar{u}_2) \in (X_2 \times \bar{\mathcal{U}}_2) \backslash (A_1 \cup A_2 \cup A_3) : \\ &\texttt{L2} \in [\texttt{L2min}, \texttt{L2max} + h(\tau_2, \texttt{MV2})], \\ &\texttt{P2} = 1,\ \texttt{L3} \geq \texttt{L3max}\} \cup \\ &\{(x_2, \bar{u}_2) \in (X_2 \times \bar{\mathcal{U}}_2) \backslash (A_1 \cup A_2 \cup A_3) : \\ &\texttt{L2} \in [\texttt{L2min}, \texttt{L2max} + h(\tau_2, \texttt{MV2})], \\ &\texttt{P2} = 1,\ \texttt{MV3} \in \{0,3\}\}, \end{aligned}$$

$$\begin{aligned} A_5 :=\{ &(x_2, \bar{u}_2) \in (X_2 \times \bar{\mathcal{U}}_2) \backslash (A_1 \cup A_2 \cup A_3) : \\ &\texttt{L2} \in [\texttt{L2min}, \texttt{L2max} + h(\tau_2, \texttt{MV2})], \\ &\texttt{P2} = 0,\ \texttt{L3} = \texttt{L3min}\} \cup \\ &\{(x_2, \bar{u}_2) \in (X_2 \times \bar{\mathcal{U}}_2) \backslash (A_1 \cup A_2 \cup A_3) : \\ &\texttt{L2} \in [\texttt{L2min}, \texttt{L2max} + h(\tau_2, \texttt{MV2})], \\ &\texttt{P2} = 0,\ \texttt{MV3} \in \{1,2\}\}, \end{aligned}$$

$$\begin{aligned} A_6 :=\{ &(x_2, \bar{u}_2) \in (X_2 \times \bar{\mathcal{U}}_2) \backslash (A_1 \cup A_2 \cup A_3) : \\ &\texttt{L2} = \texttt{L2min},\ \texttt{P2} = 1\}. \end{aligned}$$

Note that, for each $(x_2, \bar{u}_2) \in A_1$, we have

$$G_2(x_2, \bar{u}_2) \in [\texttt{L2} \ \ 0 \ \ 3 \ \ \{0,1\}]^\top.$$

Hence,

$$\begin{aligned} B(G_2(x_2, \bar{u}_2)) = (&\texttt{L2} - \texttt{L2min}) * \\ &(\texttt{L2} - \texttt{L2max} - \chi(\texttt{L2}) * h(0,3)) \\ &\leq 0. \end{aligned}$$

The latter inequality is true since

$$\texttt{L2} \geq \texttt{L2max} \quad \forall (x_2, \bar{u}_2) \in A_1,$$

and

$$h(\tau_2, 1) \leq h(0,3) \quad \forall \tau_2 \in [0, T_2].$$

Similarly, for each $(x_2, \bar{u}_2) \in A_2$, we have

$$G_2(x_2, \bar{u}_2) = [\texttt{L2} \ 0 \ 2 \ \{0,1\}]^\top.$$

Hence,

$$\begin{aligned} B(G_2(x_2, \bar{u}_2)) = (&\texttt{L2} - \texttt{L2min}) * \\ &(\texttt{L2} - \texttt{L2max} - \chi(\texttt{L2}) * h(0,2)) \\ &\leq 0. \end{aligned}$$

The latter inequality is true since $h(0,2) \geq 0$ and

$$\texttt{L2} \leq \texttt{L2min} \quad \forall (x_2, \bar{u}_2) \in A_2.$$

Now, for each $(x_2, \bar{u}_2) \in A_3$,

$$G_2(x_2, \bar{u}_2) \in [\texttt{L2} \ 0 \ \alpha(\texttt{MV2}) \ \{0,1\}]^\top,$$

where $\alpha(3) := 0$ and $\alpha(2) := 1$. Hence,

$$\begin{aligned} B(G_2(x_2, \bar{u}_2)) = (&\texttt{L2} - \texttt{L2min}) \times \\ &(\texttt{L2} - \texttt{L2max} - \chi(\texttt{L2}) * h(0, \alpha(\texttt{MV2}))) \leq 0. \end{aligned}$$

The latter inequality is true since

$$h(7,3) \leq h(0,0), \quad h(7,2) \leq h(0,1).$$

Next, for each $(x_2, \bar{u}_2) \in A_4 \cup A_5$, we have

$$G_2(x_2, \bar{u}_2) \in [\texttt{L2} \ \tau_2 \ \texttt{MV2} \ \{0,1\}]^\top.$$

Hence,

$$\begin{aligned} B(G_2(x_2, \bar{u}_2)) = (&\texttt{L2} - \texttt{L2min}) \times \\ &(\texttt{L2} - \texttt{L2max} - \chi(\texttt{L2}) * h(\tau_2 \ \texttt{MV2})) \leq 0. \end{aligned}$$

The latter inequality is true since $(A_4 \cup A_5) \subset K_e$. Hence, we conclude that (12) is satisfied. Moreover, to show (13), we note that $\tilde{C}_2 \cup \tilde{D}_2 = X_2 \times \bar{\mathcal{U}}_2$ and $G_2(x_2, \bar{u}_2) \in X_2$ for all $(x_2, \bar{u}_2) \in \tilde{D}_2$.

Next, to verify (11), we start noting that the set $U(\partial K_e) \setminus K_e$ satisfies

$$U(\partial K_e) \setminus K_e = B_1 \cup B_2,$$

where

$$B_1 := \big\{ (x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \texttt{L2} \in (\texttt{L2min} - \epsilon, \texttt{L2min}) \big\}$$

and

$$\begin{aligned} B_2 := \big\{ &(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \\ &\texttt{L2} \in (\texttt{L2max} + h(\tau_2, \texttt{MV2}), \texttt{L2max} + h(\tau_2, \texttt{MV2}) + \epsilon) \big\}. \end{aligned}$$

Furthermore, to compute the set $U(\partial K_e)\backslash K_e \cap \tilde{C}_2$, we compute the set $\tilde{C}_2$ using the following Lemma.

Lemma 1. The flow set $\tilde{C}_2$ is given by

$$\tilde{C}_2 = [C_2 \cap K_1] \cup [D_{23} \cap K_2],$$

where

$$K_1 := \mathrm{cl}((X_2 \times \bar{\mathcal{U}}_2)\backslash \tilde{D}_{26}), \quad K_2 := \mathrm{cl}((X_2 \times \bar{\mathcal{U}}_2)\backslash \tilde{I}),$$

$$\tilde{I} := \tilde{D}_{26} \cup \{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{P2} = 0,\ \mathtt{L2} \geq \mathtt{L2min}\},$$

$$\tilde{D}_{26} := \{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2} \leq \mathtt{L2min},\ \mathtt{P2} = 1\},$$

$$\begin{aligned} D_{23} := \{&(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L3} \leq \mathtt{L3min},\ \mathtt{P2} = 0\} \cup \\ &\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{MV3} \in \{1,2\},\ \mathtt{P2} = 0\}. \end{aligned}$$

□

Proof. Note that the sets D_2 and $\tilde{D}_2$ can be expressed as follows:

$$\begin{aligned} D_2 &:= D_{21} \cup D_{22} \cup D_{23} \cup D_{24} \cup D_{25}, \\ \tilde{D}_2 &:= D_{21} \cup D_{22} \cup \tilde{D}_{23} \cup D_{24} \cup D_{25} \cup \tilde{D}_{26}, \end{aligned}$$

with

$$\begin{aligned} D_{21} &:= \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2} \geq \mathtt{L2max},\ \mathtt{MV2} = 1\big\} \\ D_{22} &:= \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \tau_2 \geq T_2,\ \mathtt{MV2} \in \{2,3\}\big\} \\ D_{24} &:= \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2} \leq \mathtt{L2min},\ \mathtt{MV2} = 0\big\} \\ D_{25} &:= \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L3} \geq \mathtt{L3max},\ \mathtt{P2} = 1\big\} \cup \\ &\quad \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{MV3} \in \{0,3\},\ \mathtt{P2} = 1\big\}. \\ \tilde{D}_{23} &:= \{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L3} \leq \mathtt{L3min},\ \mathtt{P2} = 0,\ \mathtt{L2} \geq \mathtt{L2min}\} \cup \\ &\quad \{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{MV3} \in \{1,2\},\ \mathtt{P2} = 0,\ \mathtt{L2} \geq \mathtt{L2min}\}. \end{aligned}$$

Furthermore, note that $\tilde{D}_{23} = D_{23} \cap I$, where

$$I := \{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{P2} = 0,\ \mathtt{L2} \geq \mathtt{L2min}\}.$$

Hence, we have that

$$\begin{aligned} \tilde{D}_2 =& D_{21} \cup D_{22} \cup (D_{23} \cap I) \cup D_{24} \cup D_{25} \cup \tilde{D}_{26} \\ =& [D_{21} \cup D_{22} \cup D_{24} \cup D_{25}] \cup (D_{23} \cap I) \cup \tilde{D}_{26} \\ =& \{(D_{21} \cup D_{22} \cup D_{24} \cup D_{25} \cup D_{23}) \cap \\ & ([D_{21} \cup D_{22} \cup D_{24} \cup D_{25}] \cup I)\} \cup \tilde{D}_{26} \\ =& \{D_2 \cap ([D_2 \backslash D_{23}] \cup I)\} \cup \tilde{D}_{26} \\ =& (D_2 \cup \tilde{D}_{26}) \cap (([D_2 \backslash D_{23}] \cup I) \cup \tilde{D}_{26}) \\ =& (D_2 \cup \tilde{D}_{26}) \cap ([D_2 \backslash D_{23}] \cup I \cup \tilde{D}_{26}) \\ =& (D_2 \cup \tilde{D}_{26}) \cap ([D_2 \backslash D_{23}] \cup \tilde{I}). \end{aligned}$$

Next, we note that

$$\tilde{I} := \tilde{D}_{26} \cup \{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{P2} = 0,\ \mathtt{L2} \geq \mathtt{L2min}\}.$$

Hence,

$$\begin{aligned} \tilde{C}_2 =& \mathrm{cl}[(X_2 \times \bar{\mathcal{U}}_2)\backslash \tilde{D}_2] \\ =& \mathrm{cl}[(X_2 \times \bar{\mathcal{U}}_2)\backslash((D_2 \cup \tilde{D}_{26}) \cap ([D_2 \backslash D_{23}] \cup \tilde{I}))] \\ =& (C_2 \cap \mathrm{cl}[(X_2 \times \bar{\mathcal{U}}_2)\backslash \tilde{D}_{26}]) \cup \\ & \{(D_{23} \cup C_2) \cap \mathrm{cl}[(X_2 \times \bar{\mathcal{U}}_2)\backslash \tilde{I}]\} \\ =& (C_2 \cap \mathrm{cl}[(X_2 \times \bar{\mathcal{U}}_2)\backslash \tilde{D}_{26}]) \cup \{D_{23} \cap \mathrm{cl}[(X_2 \times \bar{\mathcal{U}}_2)\backslash \tilde{I}]\} \\ =& (C_2 \cap K_1) \cup \{D_{23} \cap K_2\}. \end{aligned}$$

■

Note that

$$K_1 = K_2 \cup \{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2} \geq \mathtt{L2min}, \mathtt{P2} = 0\},$$

and

$$\begin{aligned} K_2 =& \{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2} \geq \mathtt{L2min}, \mathtt{P2} = 1\} \cup \\ & \{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2} \leq \mathtt{L2min}, \mathtt{P2} = 0\}. \end{aligned}$$

Hence,

$$\begin{aligned} D_{23} \cap K_2 :=& \{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \\ & \mathtt{L3} \leq \mathtt{L3min},\ \mathtt{P2} = 0,\ \mathtt{L2} \leq \mathtt{L2min}\} \cup \\ & \{(x_2, u_2) \in X_2 \times \mathcal{U}_2 : \\ & \mathtt{MV3} \in \{1,2\},\ \mathtt{P2} = 0,\ \mathtt{L2} \leq \mathtt{L2min}\}. \end{aligned}$$

Next, we compute the set $C_2 \cap K_1$, which can be expressed as

$$C_2 \cap K_1 = \bigcup_{i=1}^{8} \tilde{C}_{2i},$$

where

$$\tilde{C}_{2i} := \tilde{C}_{2i}^a \cup \tilde{C}_{2i}^b \qquad \forall i \in \{2,4,6,8\},$$

and

$$\begin{aligned} \tilde{C}_{21} :=& \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2} \leq \mathtt{L2max},\ \mathtt{MV2} = \{1,2,3\}, \\ & \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\}, \\ \tilde{C}_{22}^a :=& \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2} \leq \mathtt{L2max},\ \mathtt{MV2} = \{1,2,3\}, \\ & \mathtt{L3} \leq \mathtt{L3max},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\}, \\ \tilde{C}_{22}^b :=& \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2} \in [\mathtt{L2min}, \mathtt{L2max}], \\ & \mathtt{MV2} = \{1,2,3\},\ \mathtt{L3} \leq \mathtt{L3max},\ \mathtt{P2} = \{1\}, \\ & \mathtt{MV3} = \{1,2\}\big\}, \end{aligned}$$

$$\begin{aligned} \tilde{C}_{23} :=& \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \\ & \mathtt{L2min} \leq \mathtt{L2} \leq \mathtt{L2max},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\}, \end{aligned}$$

$$\begin{aligned} \tilde{C}_{24}^a :=& \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2min} \leq \mathtt{L2} \leq \mathtt{L2max}, \\ & \mathtt{L3} \leq \mathtt{L3max},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\}, \\ \tilde{C}_{24}^b :=& \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2min} \leq \mathtt{L2} \leq \mathtt{L2max}, \\ & \mathtt{L3} \leq \mathtt{L3max},\ \mathtt{P2} = \{1\},\ \mathtt{MV3} = \{1,2\}\big\}, \end{aligned}$$

$$\begin{aligned} \tilde{C}_{25} :=& \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \\ & \mathtt{MV2} \in \{2,3\},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\}, \end{aligned}$$

$$\begin{aligned} \tilde{C}_{26}^a :=& \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{MV2} \in \{2,3\},\ \mathtt{L3} \leq \mathtt{L3max}, \\ & \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\}, \\ \tilde{C}_{26}^b :=& \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2} \geq \mathtt{L2min},\ \mathtt{MV2} \in \{2,3\}, \\ & \mathtt{L3} \leq \mathtt{L3max},\ \mathtt{P2} = \{1\},\ \mathtt{MV3} = \{1,2\}\big\}, \end{aligned}$$

$$\begin{aligned}\tilde{C}_{27} :=& \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2} \geq \mathtt{L2min},\ \mathtt{MV2} = \{0,2,3\},\\ & \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\},\end{aligned}$$

$$\begin{aligned}\tilde{C}_{28}^{a} :=& \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2} \geq \mathtt{L2min},\ \mathtt{MV2} = \{0,2,3\},\\ & \mathtt{L3} \in [\mathtt{L3min}, \mathtt{L3max}],\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\}.\\ \tilde{C}_{28}^{b} :=& \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L2} \geq \mathtt{L2min},\ \mathtt{MV2} = \{0,2,3\},\\ & \mathtt{L3} \in [\mathtt{L3min}, \mathtt{L3max}],\ \mathtt{P2} = \{1\},\ \mathtt{MV3} = \{1,2\}\big\}.\end{aligned}$$

Hence, we conclude that

$$\begin{aligned}&(U(\partial K_e) \setminus K_e) \cap \tilde{C}_2 = (\tilde{C}_{21} \cap B_1) \cup (\tilde{C}_{22}^{a} \cap B_1) \cup\\ &(\tilde{C}_{25} \cap B_1) \cup ((D_{23} \cap K_2) \cap B_1) \cup (\tilde{C}_{26} \cap B_1) \cup (\tilde{C}_{25} \cap B_2) \cup\\ &(\tilde{C}_{26} \cap B_2) \cup (\tilde{C}_{27} \cap B_2) \cup (\tilde{C}_{28} \cap B_2),\end{aligned}$$

with

$$\begin{aligned}\tilde{C}_{21} \cap B_1 = \big\{&(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 :\\ & \mathtt{L2} \in (\mathtt{L2min} - \epsilon, \mathtt{L2min}),\\ & \mathtt{MV2} = \{1,2,3\},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\},\end{aligned}$$

$$\begin{aligned}\tilde{C}_{22} \cap B_1 = \big\{&(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 :\\ & \mathtt{L2} \in (\mathtt{L2min} - \epsilon, \mathtt{L2min}),\ \mathtt{MV2} = \{1,2,3\},\\ & \mathtt{L3} \leq \mathtt{L3max},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\},\end{aligned}$$

$$\begin{aligned}\tilde{C}_{25} \cap B_1 = \big\{&(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 :\\ & \mathtt{L2} \in (\mathtt{L2min} - \epsilon, \mathtt{L2min}),\\ & \mathtt{MV2} \in \{2,3\},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\},\end{aligned}$$

$$\begin{aligned}\tilde{C}_{26} \cap B_1 = \big\{&(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 :\\ & \mathtt{L2} \in (\mathtt{L2min} - \epsilon, \mathtt{L2min}),\ \mathtt{MV2} \in \{2,3\},\\ & \mathtt{L3} \leq \mathtt{L3max},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\},\end{aligned}$$

$$\begin{aligned}((D_{23} \cap K_2) \cap B_1) = \{&(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 : \mathtt{L3} \leq \mathtt{L3min},\\ & \mathtt{P2} = 0,\ \mathtt{L2} \in (\mathtt{L2min} - \epsilon, \mathtt{L2min})\} \cup\\ & \{(x_2, u_2) \in X_2 \times \mathcal{U}_2 : \mathtt{MV3} \in \{1,2\},\\ & \mathtt{P2} = 0,\ \mathtt{L2} \in (\mathtt{L2min} - \epsilon, \mathtt{L2min})\},\end{aligned}$$

$$\begin{aligned}&\tilde{C}_{25} \cap B_2 = \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 :\\ &\mathtt{L2} \in (\mathtt{L2max} + h(\tau_2, \mathtt{MV2}), \mathtt{L2max} + h(\tau_2, \mathtt{MV2}) + \epsilon),\\ &\mathtt{MV2} \in \{2,3\},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\},\end{aligned}$$

$$\begin{aligned}&\tilde{C}_{26} \cap B_2 = \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 :\\ &\mathtt{L2} \in (\mathtt{L2max} + h(\tau_2, \mathtt{MV2}), \mathtt{L2max} + h(\tau_2, \mathtt{MV2}) + \epsilon),\\ &\mathtt{MV2} \in \{2,3\},\ \mathtt{L3} \leq \mathtt{L3max},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\} \cup\\ &\big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 :\\ &\mathtt{L2} \in (\mathtt{L2max} + h(\tau_2, \mathtt{MV2}), \mathtt{L2max} + h(\tau_2, \mathtt{MV2}) + \epsilon),\\ &\mathtt{MV2} \in \{2,3\},\ \mathtt{L3} \leq \mathtt{L3max},\ \mathtt{P2} = \{1\},\ \mathtt{MV3} = \{1,2\}\big\},\end{aligned}$$

$$\begin{aligned}&\tilde{C}_{27} \cap B_2 = \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 :\\ &\mathtt{L2} \in (\mathtt{L2max} + h(\tau_2, \mathtt{MV2}), \mathtt{L2max} + h(\tau_2, \mathtt{MV2}) + \epsilon),\\ &\mathtt{MV2} = \{0,2,3\},\ \mathtt{P2} = \{0\},\ \mathtt{MV3} = \{0,3\}\big\},\end{aligned}$$

$$\begin{aligned}&\tilde{C}_{28} \cap B_2 = \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 :\\ &\mathtt{L2} \in (\mathtt{L2max} + h(\tau_2, \mathtt{MV2}), \mathtt{L2max} + h(\tau_2, \mathtt{MV2}) + \epsilon),\\ &\mathtt{MV2} = \{0,2,3\},\ \mathtt{L3} \in [\mathtt{L3min}, \mathtt{L3max}],\ \mathtt{P2} = \{1\},\\ &\mathtt{MV3} = \{1,2\}\big\} \cup \big\{(x_2, \bar{u}_2) \in X_2 \times \bar{\mathcal{U}}_2 :\\ &\mathtt{L2} \in (\mathtt{L2max} + h(\tau_2, \mathtt{MV2}), \mathtt{L2max} + h(\tau_2, \mathtt{MV2}) + \epsilon),\\ &\mathtt{MV2} = \{0,2,3\},\ \mathtt{L3} \in [\mathtt{L3min}, \mathtt{L3max}],\ \mathtt{P2} = \{0\},\\ &\mathtt{MV3} = \{0,3\}\big\}.\end{aligned}$$

Next, we evaluate the product $\langle \nabla B(x_2), F_2(x_2, u_2)\rangle$ at each $(x_2, u_2) \in ((U(\partial K_e) \setminus K_e) \cap \tilde{C}_2) \times \{0,1\}$. Note that

$$\nabla B(x_2) = [\nabla_1 B(x_2)\ \nabla_2 B(x_2)\ \star\ \star]^\top,$$

where

$$\begin{aligned}\nabla_1 B(x_2) :=& 2\mathtt{L2} - (\mathtt{L2max} + \mathtt{L2min}) -\\ & \chi(\mathtt{L2}) * h(\tau_2, \mathtt{MV2}) -\\ & (\mathtt{L2} - \mathtt{L2min}) * h(\tau_2, \mathtt{MV2}) * \frac{\partial \chi}{\partial \mathtt{L2}}(\mathtt{L2}),\end{aligned}$$

$$\nabla_2 B(x_2) := -[\mathtt{L2} - \mathtt{L2min}] * \chi(\mathtt{L2}) * [F_{\mathtt{L2}}(2,0,1) + \sigma_h].$$

Hence,

$$\begin{aligned}&\langle \nabla B(x_2), F(x_2, u_2)\rangle = F_{\mathtt{L2}}(\mathtt{MV2}, \mathtt{P2}, \mathtt{P1}^a) *\\ &\big[2\mathtt{L2} - (\mathtt{L2min} + \mathtt{L2max}) - \chi(\mathtt{L2}) * h(\tau_2, \mathtt{MV2})\big] -\\ &F_{\mathtt{L2}}(\mathtt{MV2}, \mathtt{P2}, \mathtt{P1}^a) *\\ &(\mathtt{L2} - \mathtt{L2min}) * h(\tau_2, \mathtt{MV2}) * \frac{\partial \chi}{\partial \mathtt{L2}}(\mathtt{L2}) -\\ &[\mathtt{L2} - \mathtt{L2min}] * \chi(\mathtt{L2}) * [F_{\mathtt{L2}}(2,0,1) + \sigma_h].\end{aligned}$$

Next, we distinguish the following two situations:

1) When $(x_2, \bar{u}_2) \in (\tilde{C}_{21} \cap B_1) \cup (\tilde{C}_{22}^{a} \cap B_1) \cup (\tilde{C}_{25} \cap B_1) \cup ((D_{23} \cap K_2) \cap B_1) \cup (\tilde{C}_{26} \cap B_1)$, we conclude that $\mathtt{P2} = 0$, $\chi(\mathtt{L2}) = 0$, and $\frac{\partial \chi}{\partial \mathtt{L2}}(\mathtt{L2}) = 0$. Hence,

   $$\begin{aligned}&\langle \nabla B(x_2), F(x_2, u_2)\rangle =\\ &- F_{\mathtt{L2}}(\mathtt{MV2}, 0, \mathtt{P1}^a) * \big[\mathtt{L2min} + \mathtt{L2max} - 2\mathtt{L2}\big] \leq 0.\end{aligned}$$

   The latter inequality is true since according to (81), we have

   $$F_{\mathtt{L2}}(\mathtt{MV2}, 0, \mathtt{P1}^a) \geq 0 \quad \forall (\mathtt{MV2}, \mathtt{P1}^a) \in \{0,1,2\} \times \{0,1\}.$$

2) When $(x_2, u_2) \in (\tilde{C}_{25} \cap B_2) \cup (\tilde{C}_{26} \cap B_2) \cup (\tilde{C}_{27} \cap B_2) \cup (\tilde{C}_{28} \cap B_2)$, we conclude that $\chi(\mathtt{L2}) = 1$, and $\frac{\partial \chi}{\partial \mathtt{L2}}(\mathtt{L2}) = 0$. Hence,

   $$\begin{aligned}F_{\mathtt{L2}}(\mathtt{MV2}, \mathtt{P2}, \mathtt{P1}^a) &\leq F_{\mathtt{L2}}(2,0,1),\\ 0 \leq \mathtt{L2} - \mathtt{L2max} - h(\tau_2, \mathtt{MV3}) &\leq \epsilon.\end{aligned}$$

   Hence,

   $$\begin{aligned}&\langle \nabla B(x_2), F_2(x_2, u_2)\rangle \leq F_{\mathtt{L2}}(\mathtt{MV2}, \mathtt{P2}, \mathtt{P1}^a) *\\ &\big[\mathtt{L2} - \mathtt{L2min} + \epsilon\big] -\\ &[\mathtt{L2} - \mathtt{L2min}] * [F_{\mathtt{L2}}(2,0,1) + \sigma_h]\end{aligned}$$

and, for ϵ sufficiently small, we obtain

$$\langle \nabla B(x_2), F_2(x_2, u_2) \rangle \leq -\sigma_h[\mathtt{L2max} - \mathtt{L2min}]/2.$$

Hence,

$$\langle \nabla B(x_2), F_2(x_2, u_2) \rangle \leq 0.$$

■

Finally, When the Attacker Enforces $\mathtt{MV1}^a = 1$, in this case, the water level in the first tank will overflow. Indeed, the quantity of water added to the first tank in this case is bigger than the quantity removed, even when $\mathtt{P1}^a = 1$ and $\mathtt{MV2}^a = 1$.

When the Attacker Enforces $\mathtt{MV1}^a = 0$, in this case, starvation of the first stage occurs due to the constant demand of water when $\mathtt{P3} = 1$. However, this unsafe behavior can be avoided if

a. we use the logic in (88) governing `P3` with $\mathtt{L3o} = \mathtt{L3min}$.
b. we modify the logic in (17) governing `P2` using (91).
c. we modify the logic governing `P1` using

$$\mathtt{P1} := \begin{cases} 1 & \text{if } (\mathtt{L2} \leq \mathtt{L2min}, \mathtt{P1} = 0, \mathtt{L1} \geq \mathtt{L1min}) \vee \\ & (\mathtt{MV2} \in \{1,2\}, \mathtt{P1} = 0, \mathtt{L2} \geq \mathtt{L1min}), \\ 0 & \text{if } (\mathtt{L2} \geq \mathtt{L2max}, \mathtt{P1} = 1) \vee \\ & (\mathtt{MV2} \in \{0,3\}, \mathtt{P1} = 1) \vee (\mathtt{L2} \leq \mathtt{L1min}, \mathtt{P1} = 1). \end{cases} \tag{95}$$

If the logic governing one of the pumps (`P1,P2,P3`) is not modified according to the aforementioned items, then the corresponding stage will be subject to starvation.

When the Attacker Targets $\mathtt{MV2}^a$ and $\mathtt{MV3}^a$ Only and Enforces $\mathtt{MV2}^a = 1$ or $\mathtt{MV3}^a = 1$, in this case, the system remains safe since the pumps prevent overflows when they are closed.

When the Attacker Targets $\mathtt{MV2}^a$ Only and Enforces $\mathtt{MV2}^a = 0$, in this case, starvation in both Stages 2 and 3 occurs. However, if we use the logic in (88) governing `P3` with $\mathtt{L3o} = \mathtt{L3min}$, we avoid starvation of Stage 3. Moreover, if we additionally modify the logic in (17) governing `P2` using (91), the system becomes safe.

When the Attacker Targets $\mathtt{MV3}^a$ Only and Enforces $\mathtt{MV3}^a = 0$, in this case, starvation in Stage 3 occurs. However, if we use the logic in (88) governing `P3` with $\mathtt{L3o} = \mathtt{L3min}$, the system becomes safe.

In summary, we have shown that the original SWaT system cannot guarantee safety when the attacker compromises any of the following actuators `P1,P2,MV1,MV2,MV3`. However, we proposed a set of control logic changes to PLCs 1 and 2 and with these changes we were able to prove that the system is safe if the attacker compromises any of these actuators: `P1,P2,P3,MV2,MV3`. The only time SWaT cannot guarantee safety is when the attacker compromises `MV1`. The reason for this is that the amount of water coming into the first tank is higher than the amount that can be taken out by `P1`. To guarantee safety against a compromise of `MV1` we would need a physical redesign of the system so that the rate of flow of entering water is the same as the rate of flow that `P1` can take out of the first tank.

F. Discrete States from PLC Code

Programmable Logic Controllers, also known as PLCs, are industrial computers with multiple hardware modules to measure and control physical systems. PLCs rely on robust hardware that allows them to operate under extreme conditions of temperature, moisture, etc. Engineers program multiple routines that mandate how the PLC response to different states of the system. The IEC-61131-3 standard [34] groups four different languages that are broadly used by different vendors to code the routines, the most popular are Ladder Logic and Structured Text. The group of programs and routines that run on the PLCs are denoted as *control logic*.

Algorithm 1: Extract control conditions

```
Data: A PLC source code SC
Input: A set of actuators A
Result: Returns a dictionary with control
        conditions for each actuator state.
CFG:= buildCFG(SC) ;    // Build CFG from PLC source code
for a ∈ A do
    d:= SC.GetDefinitions(a);
    D.push(d);
R:= hashmap() ;      // Condition set
for d ∈ D do
    b:= CFG.GetBlockIndex(d) ;  // Get block ID for definition d
    p:= CFG.Path(b);
    s:= SC.symbExec(p) ;        // Get symbolic path condition for path p
    R.update(d) := s;
return R
```

PLCs operate reliably following a cyclic pattern called scan cycle. Roughly speaking, (1) stores data from sensor modules to a local buffer, (2) updates the network modules with local buffers, (3) runs the control logic, (4) updates local output buffers to actuator modules, (5) executes internal safety checks, and (6) repeats the cycle.

Consider the slice of PLC code shown in Listing 2. It encodes the control logic to manage a motor valve called `MV2`. We split the code into basic blocks **`B1, B2, ..., B16`** of continuous statements to explain the logic behind the PLC program.

The program starts collecting data from sensors and saving into local variables (`B1`). PLC programs use special functions like latches, counters and timers, for example in `B3`, the code employs a latch (`MV2.SR`) to collect the status of the `L2` level, the function `SETD` evaluates the inputs of the latch and updates the output according to, if the input `MV2.SR.S` is True the output `MV2.SR.Out` sets True, otherwise, if the input `MV2.SR.R` is True the output sets False. Timers use the `TONR` function. When the code calls `TONR`, the PLC evaluates if the timer is enabled (`Enable` input), if so, an internal timer is set to the `PRE` input value (in milliseconds), this timer is independent of the scan cycle to enforce real-time responses. The program has two timers (`B4`), one for each transition of `MV2 OFF` to ON and vice-versa. Block 9-14 trigger changes in MV2, being only B9 the block that turns ON MV2, while the others turn it OFF. Finally, B16 updates the actuator signals from values in local variables.

```
 1 (*Read from sensors*)
 2 'B1:'  MV2.ZSC    :=  DigitalInput (1);
 3        MV2.ZSO    :=  DigitalInput (2);
 4        L2.Level   :=  AnalogInput(1);
 5 (*Constants*)
 6 'B2:'  L2min  := 800;
 7        L2max  := 1000;
 8        T2     := 7;
 9 'B3:'  MV2.SR.Enable  :=  1;
10        MV2.SR.S       := L2.Level < L2min;
11        MV2.SR.R       := L2.Level > L2max;
12        SETD(MV2.SR);
13        MV2.Auto       := MV2.SR.Out;
14 (*Timers*)
15 'B4:'  MV2.Close_TM.PRE := T2*1000;
16        MV2.Close_TM.Enable := MV2.Cmd_Close;
17        TONR(MV2.Close_TM);
18        MV2.Open_TM.PRE := T2*1000;
19        MV2.Open_TM.Enable := MV2.Cmd_Open;
20        TONR(MV2.Open_TM);
21 'B5:'  IF MV2.ZSC THEN
22           MV2.Status  :=1;          (*MV2 fully closed*)
23 'B6:'  ELSE IF MV2.ZSO THEN
24           MV2.Status  :=2;          (*MV2 fully open*)
25 'B7:'  ELSE
26           MV2.Status  :=0;          (*MV2 in transition *)
27 'B8:'  END IF;
28 'B9:'  IF  MV2.Auto AND (NOT MV2.FC
29          AND NOT MV2.FO) THEN
30           MV2.Cmd_Close := 0;
31           MV2.Cmd_Open := 1;   (*Turn ON MV2*)
32 'B10:'    IF MV2.Open_TM.DN AND
33             (NOT MV2.ZSO) THEN
34              MV2.FC     :=  0;
35              MV2.FO     :=  1;
36 'B11:'    END IF;
37 'B12:' ELSE
38           MV2.Cmd_Close := 1;   (*Turn OFF MV2*)
39           MV2.Cmd_Open := 0;
40 'B13:'    IF MV2.Close_TM.DN
41             AND (NOT MV2.ZSC) THEN
42              MV2.FC     :=  1;
43              MV2.FO     :=  0;
44 'B14:'    END IF;
45 'B15:' END IF;
46 (*Write to actuators*)
47 'B16:'  DigitalOutput (1)  := MV2.Cmd_Close
48         DigitalOutput (2)  := MV2.Cmd_Open
```

Listing 2: Slice of PLC program controlling MV2.

We developed Algorithm 1 to analyze PLC source code, and automatically produce the set of control conditions that trigger changes in the actuator under analysis (MV2; e.g., Listing 2). To design Algorithm 1 we leveraged static program analysis concepts, like control flow analysis, symbolic execution and taint analysis. As a descriptive example, let us apply the algorithm to deduce the conditions that make MV2 to open.

To open MV2 the DigitalOutput(2) must be set to 1 (line 45). First, the algorithm builds the Control Flow Graph (CFG) depicted in Fig. 12. Nodes represent basic blocks[2] and edges show the execution order. Then, the algorithm finds the definitions, two in this case for MV2.Cmd_Open (lines 30 and 37), but only the path B9-B15-B16 matches the desired output (see Fig. 12). The symbolic execution engine will produce two conditions that satisfy the path. (1) MV2.Auto=1 $\wedge$ MV2.TON_Open_TM.DN=0, and (2) MV2.Auto=1 $\wedge$ MV2.TON_Open_TM.DN=1 $\wedge$ MV2.ZSO=1. The dependency analysis shows a link between lines 20 and 31. MV2.Open_TM.DN=1 is equivalent to '*MV2.Cmd_Open=1 for at least T2 seconds*' in B4 or ($\tau_2 \geq T_2$, MV2 =T$\uparrow$). Similarly, MV2.Auto is defined in line 13, then after processing the latch function we get that MV2.Auto is equivalent to L2.Level < L2min.

Algorithm 1 refers to these conditions as symbolic path conditions and they are stored in the R hashmap (line 10, Algorithm 1).

First two expressions in equation (18) (next subsection) describe the symbolic path conditions detailed above. The rest of the control conditions can be automatically extracted from the PLC source code using the Algorithm 1.

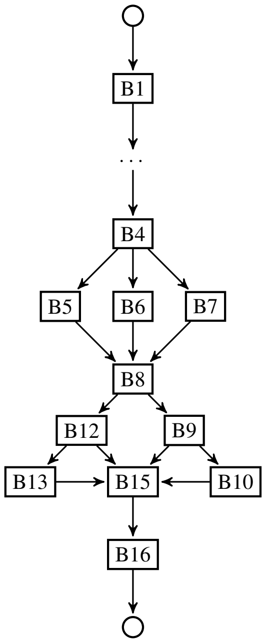


Figure 12: CFG of PLC code controlling MV2 (Listing 2).

Our approach is general and can be applied to various other control systems. For example, a PLC that manages the thermostat behavior described earlier can follow the logic described in (45), the source code written in Structured Text (see Listing 3) follows the same behavior, line 10 turns OFF the heater while line 13 turns it ON.

Our algorithm to extract the control logic from PLC source code will be made publicly available for other researchers. We believe this tool can help with the analysis of other industrial control processes, as our symbolic path conditions are generic and applicable to all PLC programs written in the "structured text" programming language.

[2] A basic block in static program analysis refers to a group of statements that run in sequence (without branches).

```
(*Read from sensors*)
'B1:'    $z :=$ AnalogInput(1);
(*Read from SCADA*)
'B2:'    $z' :=$ GetTemperature();
         $d :=$ GetTempTolerance();
(*Constants*)
'B3:'    $z_{min} := z' - d$;   (*Minimum temperature*)
         $z_{max} := z' + d$;   (*Maximum temperature*)
'B4:'    IF $z >= z_{max}$ THEN
             $q := 0$;        (*Turn OFF heater*)
'B5:'    END IF;
'B6:'    IF  $z <= z_{min}$ THEN
             $q := 1$;   (*Turn ON heater*)
'B7:'    END IF; (*Write to actuators*)
'B8:'    DigitalOutput(1) := q
```

Listing 3: PLC program of the thermostat example.

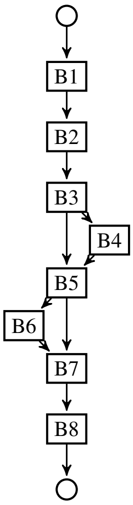


Figure 13: CFG of PLC code controlling the thermostat example (Listing 3).

G. Numerical Values for the Model

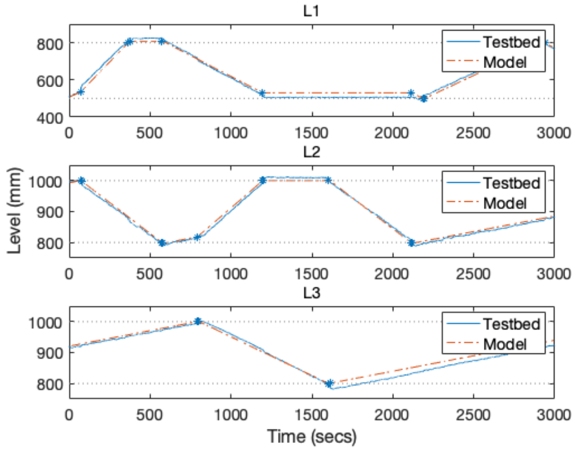


Figure 14: Nominal behavior of the system. Stars show jumps due to actuator transitions.

In this section, we provide the numerical values of the different parameters used in the modeling and the analysis of the CPS use-case. That is, using the extraction algorithms, we conclude that $(\texttt{L1min}, \texttt{L1max}) := (500, 800)$, $(\texttt{L2min}, \texttt{L2max}) := (800, 1000)$, $(\texttt{L3min}, \texttt{L3max}) := (800, 1000)$, and $(T_1, T_2, T_3) := (9, 7, 7)$. Furthermore, the numerical values of the rate of change of the water levels $\texttt{L1}$, $\texttt{L2}$, and $\texttt{L3}$; namely, F_{L1}, F_{L2}, and F_{L3}, have been computed using linear regression.

1) **In the absence of attacks and when P3 is always ON.**
   - $F_{\mathrm{L3}} \equiv F_{\mathrm{L3}}(\texttt{MV3})$ with $F_{\mathrm{L3}}(0) = -0.15$, $F_{\mathrm{L3}}(1) = 0.16$, $F_{\mathrm{L3}}(2) = F_{\mathrm{L3}}(3) = 0.11$.
   - $F_{\mathrm{L2}} \equiv F_{\mathrm{L2}}(\texttt{MV2}, \texttt{P2})$ with
   $$F_{\mathrm{L2}}(0,0) = 0,\ F_{\mathrm{L2}}(1,0) = 0.46,\ F_{\mathrm{L2}}(1,1) = 0.13,$$
   $$F_{\mathrm{L2}}(2,1) = F_{\mathrm{L2}}(3,1) = -0.28,$$
   $$F_{\mathrm{L2}}(2,0) = F_{\mathrm{L2}}(3,0) = 0.29.$$
   - $F_{\mathrm{L1}} \equiv F_{\mathrm{L1}}(\texttt{MV1}, \texttt{P1})$ with
   $$F_{\mathrm{L1}}(0,1) = -0.45,\ F_{\mathrm{L1}}(1,1) = 0.41,$$
   $$F_{\mathrm{L1}}(2,1) = F_{\mathrm{L1}}(3,1) = -0.15,$$
   $$F_{\mathrm{L1}}(0,0) = 0,\ F_{\mathrm{L1}}(1,0) = 0.9,$$
   $$F_{\mathrm{L1}}(2,0) = F_{\mathrm{L1}}(3,0) = 0.81.$$

2) **When only $\texttt{P2}^a$ is attacked and P3 is always ON.**
   - $F_{\mathrm{L3}} \equiv F_{\mathrm{L3}}(\texttt{MV3}, \texttt{P2}^a)$ with
   $$F_{\mathrm{L3}}(\texttt{MV3}, 0) = -0.15 \quad \forall \texttt{MV3} \in \{0,1,2,3\},$$
   and $F_{\mathrm{L3}}(\texttt{MV3}, 1) \equiv F_{\mathrm{L3}}(\texttt{MV3})$ as in 1).
   - $F_{\mathrm{L2}} \equiv F_{\mathrm{L2}}(\texttt{MV2}, \texttt{MV3}, \texttt{P2}^a)$ with
   $$F_{\mathrm{L2}}(\texttt{MV2}, 1, \texttt{P2}^a) \equiv F_{\mathrm{L2}}(\texttt{MV2}, \texttt{P2}^a),$$
   $$F_{\mathrm{L2}}(\texttt{MV2}, \texttt{MV3}, 0) \equiv F_{\mathrm{L2}}(\texttt{MV2}, 0)$$
   $$F_{\mathrm{L2}}(\texttt{MV2}, 0, \texttt{P2}^a) \equiv F_{\mathrm{L2}}(\texttt{MV2}, 0)$$
   $$F_{\mathrm{L2}}(\texttt{MV2}, 2, 1) \equiv F_{\mathrm{L2}}(\texttt{MV2}, 1)$$
   as in 1). Furthermore,
   $$F_{\mathrm{L2}}(\texttt{MV2}, 2, \texttt{P2}^a) \equiv F_{\mathrm{L2}}(\texttt{MV2}, 3, \texttt{P2}^a).$$
   - $F_{\mathrm{L1}} \equiv F_{\mathrm{L1}}(\texttt{MV1}, \texttt{P1})$ as in 1).

3) **In the absence of attacks and P3 is not always ON.**
   - $F_{\mathrm{L3}} \equiv F_{\mathrm{L3}}(\texttt{MV3}, \texttt{P3})$ with $F_{\mathrm{L3}}(\texttt{MV3}, 1)$ as in 1) and $F_{\mathrm{L3}}(0,0) = 0$, $F_{\mathrm{L3}}(1,0) = 0.36$, $F_{\mathrm{L3}}(2,0) = F_{\mathrm{L3}}(3,0) = 0.3$.
   - $F_{\mathrm{L2}} \equiv F_{\mathrm{L2}}(\texttt{MV2}, \texttt{P2})$ as in 1).
   - $F_{\mathrm{L1}} \equiv F_{\mathrm{L1}}(\texttt{MV1}, \texttt{P1})$ as in 1).

4) **When only $\texttt{P2}^a$ is attacked and P3 is not always ON.**
   - $F_{\mathrm{L3}} \equiv F_{\mathrm{L3}}(\texttt{MV3}, \texttt{P3}, \texttt{P2}^a)$ with
   $$F_{\mathrm{L3}}(\texttt{MV3}, \texttt{P3}, 1) \equiv F_{\mathrm{L3}}(\texttt{MV3}, \texttt{P3})$$
   $$F_{\mathrm{L3}}(\texttt{MV3}, \texttt{P3}, 0) \equiv F_{\mathrm{L3}}(0, \texttt{P3})$$
   as in 3).
   - $F_{\mathrm{L2}} \equiv F_{\mathrm{L2}}(\texttt{MV2}, \texttt{MV3}, \texttt{P2}^a)$ as in 2).
   - $F_{\mathrm{L1}} \equiv F_{\mathrm{L1}}(\texttt{MV1}, \texttt{P1})$ as in 1).

5) **When $\texttt{P1}^a$ in Attacked and P3 is not Always ON.**
   - $F_{\mathrm{L3}} \equiv F_{\mathrm{L3}}(\texttt{MV3}, \texttt{P3})$ as in 3).
   - $F_{\mathrm{L2}} \equiv F_{\mathrm{L2}}(\texttt{MV2}, \texttt{P2}, \texttt{P1}^a)$ with
   $$F_{\mathrm{L2}}(\texttt{MV2}, \texttt{P2}, 1) \equiv F_{\mathrm{L2}}(\texttt{MV2}, \texttt{P2})$$

$$F_{\mathrm{L2}}(\mathtt{MV2}, \mathtt{P2}, 0) \equiv F_{\mathrm{L2}}(0, \mathtt{P2})$$

as in 1).

- $F_{\mathrm{L1}} \equiv F_{\mathrm{L1}}(\mathtt{MV1}, \mathtt{MV2}, \mathtt{P1}^a)$ with

$$F_{\mathrm{L1}}(\mathtt{MV1}, \mathtt{MV2}, 0) = F_{\mathrm{L1}}(\mathtt{MV1}, 0, \mathtt{P1}^a) \equiv F_{\mathrm{L1}}(\mathtt{MV1}, 0)$$

$$F_{\mathrm{L1}}(\mathtt{MV1}, 1, 1) = F_{\mathrm{L1}}(\mathtt{MV1}, 2, 1) \equiv F_{\mathrm{L1}}(\mathtt{MV1}, 1)$$

as in 1).

To test if our model follows the dynamics of the real-world system, we implemented our equations in Matlab's HyEq toolbox [51]. Figure 14 and Figure 15 shows traces of the real-world system labeled as **Testbed** of 3000 *seconds* under nominal conditions. It also shows the execution of the **Model** during the same period. The model follows the behavior of the Testbed, including triggering actuator transitions synchronously. The different initial conditions of the state vector x are:

$$\begin{array}{llll} \mathtt{L1} = 506.6; & \tau 1 = 1; & \mathtt{MV1} = \mathrm{ON}; & \mathtt{P1} = \mathrm{ON}; \\ \mathtt{L2} = 992.7; & \tau 2 = 1; & \mathtt{MV2} = \mathrm{ON}; & \mathtt{P2} = \mathrm{ON}; \\ \mathtt{L3} = 920.8; & \tau 3 = 1; & \mathtt{MV3} = \mathrm{ON}; & \mathtt{P3} = \mathrm{ON}. \end{array}$$

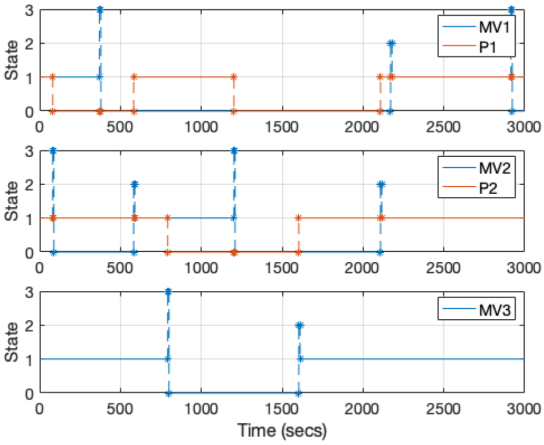


Figure 15: Nominal behavior of the system. Stars show jumps due to actuator transitions.

Figure 16 shows the behavior of our model under multiple operational scenarios of the SWaT testbed, as another example of the fidelity of our model.

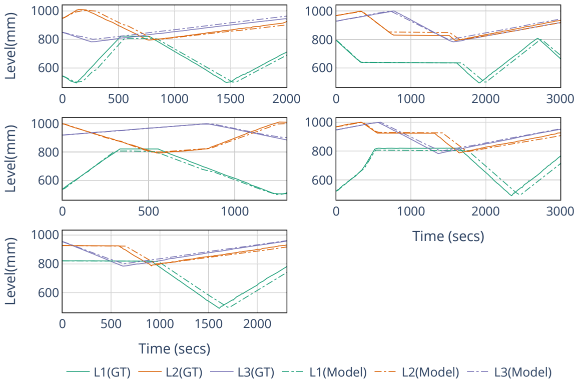


Figure 16: Comparison of multiple traces of the model against ground truth (GT)