Proceedings of Machine Learning Research vol 144:1–19, 2021 3rd Annual Conference on Learning for Dynamics and Control

Learning-based attacks in Cyber-Physical Systems: Exploration, Detection, and Control Cost trade-offs

Anshuka Rangi* ARANGI@UCSD.EDU
Department of Electrical and Computer Engineering, University of California San Diego, CA 92093

Mohammad Javad Khojasteh* MKHOJAST@MIT.EDU
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, MA, 02139

Massimo Franceschetti MFRANCESCHETTI@UCSD.EDU
Department of Electrical and Computer Engineering, University of California San Diego, CA 92093

Editors: A. Jadbabaie, J. Lygeros, G. J. Pappas, P. A. Parrilo, B. Recht, C. J. Tomlin, M. N. Zeilinger

Abstract

We study the problem of learning-based attacks in linear systems, where the communication channel between the controller and the plant can be hijacked by a malicious attacker. We assume the attacker learns the dynamics of the system from observations, then overrides the controller's actuation signal, while mimicking legitimate operation by providing fictitious feedback about the sensor readings to the controller. On the other hand, the controller is on a lookout to detect the presence of the attacker and tries to enhance the detection performance by carefully crafting its control signals. We study the trade-offs between the information acquired by the attacker from observations, the detection capabilities of the controller, and the control cost. Specifically, we provide tight upper and lower bounds on the expected ϵ-deception time, namely the time required by the controller to make a decision regarding the presence of an attacker with confidence at least $(1-\epsilon\log(1/\epsilon))$. We then show a probabilistic lower bound on the time that must be spent by the attacker learning the system, in order for the controller to have a given expected ϵ-deception time. We show that this bound is also order optimal, in the sense that if the attacker satisfies it, then there exists a learning algorithm with the given order expected deception time. Finally, we show a lower bound on the expected energy expenditure required to guarantee detection with confidence at least $1-\epsilon\log(1/\epsilon)$.
Keywords: Data poisoning attack, Man in the middle attack, Cyber physical system

1. Introduction

Attacks directed to Cyber-Physical Systems (CPS) can have catastrophic consequences ranging from hampering the economy through financial scams, to possible losses of human lives through hijacking autonomous vehicles and drones, see Pasqualetti et al. (2013); Shoukry et al. (2018); Hoehn and Zhang (2016). In this framework, two important problems arise: understanding of the regime where the system can be attacked, and designing ways to mitigate these attacks and render the system secure, see Ma et al. (2019); Zhang et al. (2020b); Jun et al. (2018); Zhan et al. (2020); Chen and Zhu (2019); Vemprala and Kapoor (2020); Ferdowsi and Saad (2018); Mao et al. (2020); Khojasteh et al. (2019, 2021); Rangi et al. (2021). Techniques developed to secure CPS include watermarking, moving target and baiting, and typically require either a loss of performance, or

* indicates equal contribution.

additional resources available at the controller, see Satchidanandan and Kumar (2017); Mo et al. (2015); Kanellopoulos and Vamvoudakis (2019); Flamholz et al. (2019).

In this paper, we focus on the former aspect of the problem, namely understanding the regime under which the system can be attacked. We focus on linear plants and on an important and widely used class of attacks based on the “man-in-the-middle" (MITM) technique. In this case, the attacker takes over the physical plant’s control and feedback signals, and acts as a malicious controller for the plant and fictitious plant for the controller. By doing so, it overrides the control signals with malicious inputs aimed at destroying the plant; and it overrides the feedback signals to the controller, trying to mimic the safe and legitimate operation of the system. In learning based MITM attack, we assume that the attacker has full access to both sensor and control signals, but the plant dynamics are unknown to the attacker. Thus, the attacker needs to learn about the plant in order to being able to generate the fictitious signals to the controller that allow the attacker to remain undetected for the time needed to cause harm. On the other hand, the controller has perfect (or nearly perfect) knowledge of the system dynamics and is actively looking out for an anomalous behaviour in the feedback signals from the plant. This assumed information pattern is justified, since the controller is typically tuned in much longer than the attacker, and has knowledge of the system dynamics to a far greater precision than the attacker. Following the detection of the attacker, the controller can shut the plant down, or switch to a “safe” mode where the system is secured using additional resources, and the attacker is prevented from causing additional "harm" to the plant, see Dibaji et al. (2019); Weerakkody et al. (2019); Teixeira et al. (2015); Hashemi and Ruths (2020).

We consider a learning-based MITM attack that evolves in two phases: *exploration and exploitation*. In the exploration phase, the attacker observes the plant state and control inputs, and learns the plant dynamics. In the exploitation phase, the attacker hijacks the plant, and utilizes the learned estimate to feed the fictitious feedback signals to the controller. During this phase, the attacker may also refine its estimate by continuing to learn. Within this context, our results are as follows: first, we provide a lower bound on the expected ϵ-deception time, namely the time required by the controller to make a decision regarding the presence of an attacker with confidence at least $1 - \epsilon \log(1/\epsilon)$. This bound is expressed in terms of the parameters of the attacker’s learning algorithm and the controller’s strategy. Second, we show that there exists a learning-based attack and a detection strategy such that a matching upper bound on the expected ϵ-deception time is obtained. We then show that for a wide range of learning algorithms, if the expected ϵ-deception time is at least of duration D, then the duration of the exploration phase of the attacker must be at least $\Omega(D/\log(1/\epsilon))$, as $\epsilon \to 0$. We establish that this bound is also order-optimal since there exists a learning algorithm such that if the duration of the exploration phase is $O(D/\log(1/\epsilon))$ as $\epsilon \to 0$, then the expected ϵ-deception time is at least D. Finally, we show that if the controller wants to detect the attacker in at most D duration with confidence at least $1 - \epsilon \log(1/\epsilon)$, then the expected energy expenditure on the control signal must be at least of order $\Omega(D/\log(1/\epsilon))$, as $\epsilon \to 0$.

All proofs are available at Rangi et al. (2020b).

2. Related Work

There is a wide range of recent research on learning-based control for linear systems Dean et al. (2019); Sarkar and Rakhlin (2019); Berkenkamp et al. (2017); Fisac et al. (2018); Khojasteh et al. (2020); Vrabie et al. (2009); Jiang and Jiang (2012); Cheng et al. (2020); Fan et al. (2020); Lederer et al. (2019); Buisson-Fenet et al. (2020). In these works, learning algorithms are proposed to design controllers in the presence of uncertainty. In contrast, in our setting we assume that the controller has full knowledge of the system dynamics, while the attacker may take advantage of these algorithms. Thus, our focus is not on the optimal control design given the available data, but rather on the trade-offs between the attacker's learning capability, the controller's detection strategy, and the control cost.

The MITM attack has been extensively studied in control systems for two special cases, namely, the replay attack and the statistical duplicate attack. The detection of replay attacks has been studied in Mo et al. (2014, 2015); Miao et al. (2013), and ways to mitigate these attacks have been studied in Zhu and Martínez (2014). Likewise, the ways to detect and mitigate statistical duplicate attacks has been studied in Satchidanandan and Kumar (2017); Smith (2015); Porter et al. (2020). These works do not consider learning-enabled attackers, and analyze the performance of the controller for only a specific detection strategy. In contrast, we investigate learning-enabled attacks, and present trade-offs between the attacker's learning capability through observations, the controller detection strategy, and the control cost. Learning based attacks have been recently considered in Khojasteh et al. (2019, 2021); Ziemann and Sandberg (2020). In Khojasteh et al. (2019, 2021), a variance based detection strategy has been investigated to present bounds on the probabilities of detection (or false alarm) of the attacker. In Ziemann and Sandberg (2020), an optimization-based controller is proposed that has the additional capability of injecting noise to interfere with the learning process of the attacker. Here, we consider a wider class of learning-based attacks and detection strategies, and provide tight trade-offs for these attacks.

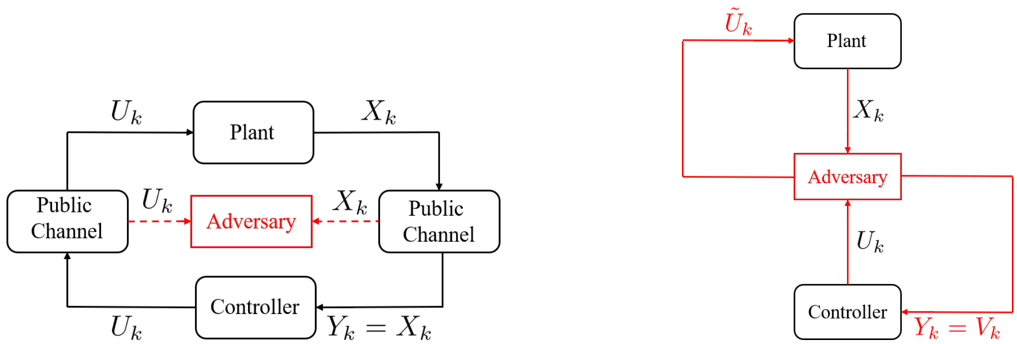
Multiple variants of MITM attacks are studied in Reinforcement Learning (RL). In Rakhsha et al. (2020), the work studies the MITM attacks under the assumption that the attacker has perfect knowledge of the underlying MDP. The results are further extended to the setting where attacker has no knowledge of the underlying MDP Rakhsha et al. (2021). This is analogues to studying learning based attacks in RL where the attacker eavesdrops on the actions performed by the learner and manipulates the feedback from the environment. In Zhang et al. (2020a), the work studies the feasibility of MITM attack under the constraint on the amount of contamination introduced by the attacker in the feedback signal. The relationship between the problem of designing optimal MITM attack in RL and the problem of designing optimal control is discussed in Zhu (2018). Finally, the learning based MITM attacks are also an active area of research in the Multi-Armed Bandits (MAB), see Jun et al. (2018); Ma et al. (2018); Bogunovic et al. (2021); Rangi et al. (2021). In the same spirit of our work, these works study the feasibility of the attacks, and provide bounds on the amount of contamination needed by the attacker to achieve its objective. However, these works do not consider the possibility of the detection of the attacker. In this work, we focus on understanding the regime where the system can be attacked without the detection of the attacker.

3. Problem Setup

We consider the networked control system depicted in Fig. 1(a) and Fig. 1(b), where the plant dynamics are described by a discrete-time and linear time-invariant (LTI) system, namely at time $k \in \mathbb{N}$, we have

$$X_{k+1} = AX_k + U_k + W_k, \tag{1}$$

where X_k, U_k, W_k are vectors of dimension $M \times 1$ representing the plant state, control input, and plant disturbance respectively, and A is a matrix of dimension $M \times M$, representing the open-



(a) Exploration Phase (b) Exploitation Phase

Figure 1: System model during the two attack phases.

loop gain of the plant. At time k, the controller observes the feedback signal Y_k and generates a control signal U_k as a function of $Y_{1:k} = \{Y_1, \ldots, Y_k\}$. The initial state X_0 is known to both the controller and the attacker, and is independent of the disturbance sequence $\{W_k\}_{k=1}^{\infty}$, where W_k is i.i.d. Gaussian noise $\mathcal{N}(0, \sigma^2 I_M)$ with PDF known to both the parties, and I_M is the identity matrix of dimension $M \times M$. Our results can also be extended to the scenario where the PDF of the noise known to the attacker is different from the actual PDF of the noise (or PDF known to the controller). With a slight loss of generality, we assume that $U_0 = W_0 = 0$ for analysis.

The controller attempts to detect the presence of the attacker based on the observations $Y_{1:k}$. When the controller detects an attack, it shuts the system down and prevents the attacker from causing further "damage" to the plant. The controller is aware of the plant dynamics in (1), and knows the gain A. This is justified because one can assume that the controller is tuned to the plant for a long duration and thus has knowledge of A to a great precision. On the other hand, the attacker only knows the form of the state evolution equation (1), but does not know the gain matrix A.

4. Learning based Attacks

We consider learning based attacks that evolve in two phases.

Phase 1: Exploration. Let L be the duration of the exploration phase. For all $k \leq L$, as illustrated in Fig. 1(a), the attacker passively eavesdrops on the control input U_k and the plant state $Y_k = X_k$ with the objective of learning the open loop gain of the plant. We let $\hat{A}_k$ be the attacker's estimate of A at time step k. The duration L can be considered as the cost incurred by the attacker, since its actions are limited to eavesdropping during this phase.

Phase 2: Exploitation. The exploration phase is followed by the exploitation phase. For all $k \geq L+1$, as illustrated in Fig. 1(b), the attacker hijacks the system and feeds a malicious control signal $\tilde{U}_k$ to the plant in order to destroy the plant. Additionally, the attacker may continue to learn about A, and utilizes its estimate $\hat{A}_k$ to design a fictitious feedback signal $Y_k = V_k$ in Fig. 1(b) to deceive the controller, namely

$$V_{k+1} = \hat{A}_k V_k + U_k + \tilde{W}_k, \tag{2}$$

where for all $k \geq L+1$, $\tilde{W}_k$ are i.i.d. with $f_{\tilde{W}} = f_W = \mathcal{N}(0, \sigma^2 I_M)$. Let R denote an attack strategy whose feedback signal satisfies (2). Thus, for all $L > 0$, our class of learning based attacks is

$$\mathcal{A}(L) = \{R : \text{for all } k \leq L, Y_k = X_k \text{ and for all } k \geq L+1, Y_k = V_k\}. \tag{3}$$

Note that in the class $\mathcal{A}(L)$, the learning of A may or may not continue during the exploitation phase. Additionally, the attacker may use different learning algorithms in the two phases.

If the attacker learns A perfectly, i.e. $\hat{A}_k = A$, then (2) will perfectly mimic the plant behavior, making it impossible for the controller to detect the attacker. Otherwise, the controller can attempt to detect the presence of the attacker by testing for statistical deviations from the typical behavior in (1). The following example illustrates this point.

Example 1 *Let $R^* \in \mathcal{A}(L)$ be an attack whose learning is only limited to the exploration phase, namely $\hat{A}_k = \hat{A}_L$ for all $k \geq L+1$. Also, let $\|\cdot\|_{op}$ be the operator norm induced by the Euclidean norm $\|\cdot\|_2$ when applied to a matrix. In the exploration phase there is no interference from the attacker and for all $k \leq L$, the observation $Y_k = X_k$ satisfies*

$$Y_{k+1} - AY_k - U_k = W_k \sim \text{ i.i.d. } f_W. \tag{4}$$

In the exploitation phase, for all $k \geq L+1$, the controller observation $Y_k = V_k$ satisfies

$$V_{k+1} - AV_k - U_k = V_{k+1} - AV_k + \hat{A}_L V_k - \hat{A}_L V_k - U_k = \tilde{W}_k + \left(\hat{A}_L - A\right) V_k, \tag{5}$$

where (5) follows from (4). Since $\tilde{W}_k$ and W_k have the same distribution and $\|Ax\|_2 \leq \|A\|_{op}\|x\|_2$ holds, the controller can test the statistical deviation of (4) from (5). In this case, the detection of the attack is controlled by two factors: the estimation error $\|\hat{A}_L - A\|_{op}$ and the fictitious signal V_k.

At the controller's side, the detection becomes easier when the error $\|\hat{A}_L - A\|_{op}$ increases. Thus, at the attacker's side it is desirable to reduce the error $\|\hat{A}_L - A\|_{op}$. This can be done by increasing the duration L, and incurring an additional learning cost.

The detection is also easier if the energy of the fictitious signal V_k is large. Since V_k is a function of the control signal U_{k-1}, it follows that the energy spent by the controller can help in the detection of the attacker.

We then conclude that the probability of successful detection (or the time required to detect the attacker with a given confidence) should reveal a trade-off between the duration L of the exploration phase (or the estimation error $\|\hat{A}_L - A\|_{op}$), and the energy of the fictitious signal (or of the control signal). In this paper we quantify both upper bound and lower bound on this trade-off.

4.1. Performance Measures

Definition 1 *The decision time τ is the time at which the controller makes a decision regarding the presence or absence of the attacker.*

Definition 2 *The probability of deception is the probability of the attacker deceiving the controller and remaining undetected at the decision time τ, namely $P^{\tau}_{\text{Dec}} \triangleq \mathbb{P}(\hat{\Theta}_\tau = 0 | \Theta_\tau = 1)$, where $\hat{\Theta}_\tau$ denotes the decision of the controller at the decision time τ, and the hijack indicator Θ_k at time k is*

$$\Theta_k \triangleq \begin{cases} 0, & \forall j \leq k: \; Y_j = X_j\,; \\ 1, & \textit{otherwise}. \end{cases} \tag{6}$$

Likewise, the probability of false alarm is the probability of detecting the attacker when it is not present at the decision time τ, namely $P^{\tau}_{\text{FA}} \triangleq \mathbb{P}(\hat{\Theta}_\tau = 1 | \Theta_\tau = 0)$.

In the class $\mathcal{A}(L)$ in (3), for all $k \leq L$, we have that $\Theta_k = 0$ (exploration phase); and for all $k \geq L+1$, we have $\Theta_k = 1$ (exploitation phase).

Definition 3 *For all attacks in the class $\mathcal{A}(L)$ and $0 < \epsilon < 1$, the ϵ-deception time $T(\epsilon)$ is the time required by the controller to make a decision, with $P^{\tau}_{\text{Dec}} \leq \epsilon \log(1/\epsilon)$, where $\tau = L + T(\epsilon) + 1$.*

Thus, $T(\epsilon)$ is the largest possible duration during which the attacker can deceive the controller, and remain undetected with confidence at least $1 - \epsilon \log(1/\epsilon)$, namely for all $L+1 \leq k \leq T(\epsilon) + L$, we have

$$\mathbb{P}(\hat{\Theta}_k = \Theta_k | \Theta_k = 1) = \mathbb{P}(\hat{\Theta}_k = 1 | \Theta_k = 1) < 1 - \epsilon \log(1/\epsilon). \tag{7}$$

Definition 4 *For all $n > L$, the expected deception cost of the attacker until time n is defined as*

$$C(n) \triangleq \frac{1}{n} \mathbb{E}\Bigg[\sum_{k=L+1}^{n} \frac{V_{k-1}^T (\hat{A}_{k-1} - A)^T (\hat{A}_{k-1} - A) V_{k-1}}{2\sigma^2} \Bigg]. \tag{8}$$

4.2. Main results

We start with defining a non-divergent learning algorithm.

Definition 5 *A learning algorithm is non-divergent if its estimation error is non-increasing in the duration of the learning, namely for all $k_2 > k_1$, we have $\|\hat{A}_{k_2} - A\|_{op} \leq \|\hat{A}_{k_1} - A\|_{op}$.*

We introduce the following notation. Let $p_0(y_{1:\tau})$ be the conditional probability of $y_{1:\tau}$ given the attacker did not hijack the system, namely $\Theta_1 = \ldots \Theta_L = \Theta_{L+1} = \ldots \Theta_\tau = 0$, where $y_{1:\tau}$ denotes the realization of the random variables $Y_1, \ldots, Y_\tau$. Likewise, let $p_1(y_{1:\tau})$ be the conditional probability of $y_{1:\tau}$ given the attacker has hijacked the system, namely $\Theta_1 = \ldots = \Theta_L = 0$ and $\Theta_{L+1} = \ldots \Theta_\tau = 1$. The following proposition characterises the KL divergence $D(p_1(Y_{1:\tau}) || p_0(Y_{1:\tau}))$ between $p_1(Y_{1:\tau})$ and $p_0(Y_{1:\tau})$, and is useful to derive our main results.

Proposition 1 *For all attacks in the class $\mathcal{A}(L)$ and $n > L$, the cumulative KL divergence is*

$$D(p_1(Y_{1:n}) || p_0(Y_{1:n})) = nC(n). \tag{9}$$

The KL divergence between the distributions p_0 and p_1 is characterized by $C(n)$, and is the key quantity to establish both the lower bound and the upper bound on $T(\epsilon)$. If the PDF of the noise known to the attacker is different from the actual PDF of the noise (or the PDF known to the controller), Proposition 1 can be modified to include this discrepancy, and an additional non-negative term would be added to $C(n)$. The bounds on $T(\epsilon)$ will follow along the same lines.

The following theorem presents a lower bound on $\mathbb{E}[T(\epsilon)]$ that holds for any detection strategy. The bound is expressed in terms of $C(n)$, which depends on the attacker's learning algorithm, the fictitious signal and the control signal in (2).

Theorem 1 *For all attacks in $\mathcal{A}(L)$ and $\tau > L$, if*

$$P^{\tau}_{\text{Dec}} = O(|\epsilon \log \epsilon|) \text{ and } P^{\tau}_{FA} = O(|\epsilon \log \epsilon|), \text{ as } \epsilon \to 0, \tag{10}$$

then the deception time $T(\epsilon) = \tau - L - 1$ is

$$\mathbb{E}[T(\epsilon)] \geq \frac{\log(1/\epsilon)}{C(n_0)} + o(\log(1/\epsilon)) \quad \text{as } \epsilon \to 0, \tag{11}$$

where $n_0 = \max\{n > L : nC(n) < \log(1/\epsilon)\}$.

It follows that for any detection strategy with probability of error $O(|\epsilon \log \epsilon|)$, the expected ϵ-deception time is at least $\Omega\left(\log(1/\epsilon)/C(n_0)\right)$. The next theorem establishes that the lower bound in Theorem 1 is tight.

Theorem 2 *There exists an attack in $\mathcal{A}(L)$ and a detection strategy such that at $\tau > L$, we have*

$$P_{\text{Dec}}^{\tau} = O(\epsilon) \text{ and } P_{FA}^{\tau} = O(\epsilon), \text{ as } \epsilon \to 0, \tag{12}$$

and the deception time $T(\epsilon) = \tau - L - 1$ is

$$\mathbb{E}[T(\epsilon)] \leq \frac{\log(1/\epsilon)}{C(n_0+1)} + o(\log(1/\epsilon)), \quad \text{as } \epsilon \to 0. \tag{13}$$

In Theorems 1 and 2, as $\epsilon \to 0$, we have that $C(n_0) \to C(n_0+1)$, and $|\epsilon| \leq |\epsilon \log \epsilon|$. Thus, the lower bound and the upper bound in Theorems 1 and 2 are tight. It turns out that the attack achieving the upper bound on $\mathbb{E}[T(\epsilon)]$ in Theorem 2 learns about A in the exploration phase only, and focuses on destabilizing the system in the exploitation phase. The corresponding detection strategy is a classic sequential probability ratio test (Wald et al. (1948)), which computes the ratio of the posterior probability of the two hypotheses, namely the attacker is present or absent, and makes a decision when this ratio crosses the threshold $\log(1/\epsilon)$. While this strategy has been previously studied under the assumption that the samples $y_{1:n}$ are identically and independently distributed (i.i.d) (Chernoff (1959); Rangi et al. (2018b,a, 2020a)), here we extend the analysis to the samples dependent on both the control input and the state of the feedback signal at the controller.

We point out that to extend these results to non-linear systems, a key step would be finding an analogue of Proposition 1 in a non-linear setting. This proposition relates the KL divergence to the expected deception cost $C(n)$, which is a function of the fictitious signal and the error in the estimation of A. For non-linear systems, an equivalent relationship needs to be derived between the KL divergence, the fictitious signal and the error in the estimation of non-linear system dynamics. The proof of the Theorems 1 and 2 can then be obtained using a similar argument, given an analogue of Proposition 1 for non-linear systems.

Next, we derive some useful implications of Theorems 1 and 2. For simplicity of presentation, in the following we restrict the class of learning algorithms in the exploration phase, although results can also be extended to more general settings.

Definition 6 *A learning algorithm is said to be convergent if there exists an $\alpha \geq 1$ such that for all $\eta > 0$ and time step $k > 0$, we have*

$$\mathbb{P}(\|\hat{A}_k - A\|_{op} > \eta) \leq \frac{c}{(\eta^2 k)^{\alpha}}. \tag{14}$$

It follows that any convergent learning algorithm provides an unbiased estimate of A as the learning time $k \to \infty$, and the operator norm of the estimation error converges to the interval $[-\eta, +\eta]$ at rate $O(1/(\eta^2 k)^{\alpha})$. There are many convergent learning algorithms. For example, for scalar systems and measurable control policy, the Least Squares (LS) algorithm in Rantzer (2018) satisfies

$$\mathbb{P}(|\hat{A}_k - a| > \eta) \leq \frac{2}{(1+\eta^2)^{k/2}}. \tag{15}$$

For the vector case sufficiently large learning time k, if the control input is $U_k = -\bar{K} X_k$ and $A - \bar{K}$ is a marginally stable matrix, then the LS algorithm in Simchowitz et al. (2018) satisfies

$$\mathbb{P}(\|\hat{A}_k - A\|_{op} > \eta) \leq \frac{c_1}{e^{\eta^2 k}}, \tag{16}$$

where $c_1 > 0$ is a constant.

The following theorem provides a lower bound on the duration of the exploration phase for the attacker to achieve a given expected ϵ-deception time.

Theorem 3 *For all $0 < \delta < 1$ and $D > 0$, and all attacks in $\mathcal{A}(L)$ using a convergent learning algorithm in the exploration phase and a non-divergent learning algorithm in the exploitation phase, if $\mathbb{E}[T(\epsilon)] \geq D + o(1)$ as $\epsilon \to 0$, then with probability at least $1 - \delta$ the following asymptotic inequality holds*

$$L \geq \frac{D\tilde{C}(n_0)}{\log(1/\epsilon)} \left(\frac{c}{\delta}\right)^{1/\alpha} + o\left(\frac{1}{\log(1/\epsilon)}\right), \text{ as } \epsilon \to 0, \tag{17}$$

where $\tilde{C}(n) = \mathbb{E}\big[\sum_{k=L+1}^{n} V_{k-1}^T V_{k-1}\big]/(2\sigma^2 n)$.

The following theorem establishes that the lower bound on L in Theorem 3 is order optimal, and a matching order-optimal bound on L holds for the LS algorithm in Simchowitz et al. (2018).

Theorem 4 *For all $0 < \delta < 1$ and $D > 0$, using the LS algorithm in Simchowitz et al. (2018) in the exploration phase only, and assuming the control input is $U_k = -\bar{K} X_k$, where $A - \bar{K}$ is a marginally stable matrix, if*

$$L = D\tilde{C}(n_0) \log(c_1/\delta)/\log(1/\epsilon) + o(1/\log(1/\epsilon)) \text{ as } \epsilon \to 0, \tag{18}$$

then, with probability at least $1 - \delta$ we have

$$\mathbb{E}[T(\epsilon)] \geq D + o(1), \text{ as } \epsilon \to 0. \tag{19}$$

The choice of the control policy can play a crucial role in the reduction of the deception time. However, this can occur at the expense of the energy used to construct the control signal U_k. The following theorem provides a lower bound on the amount of energy that the controller needs to spend to achieve a desired expected ϵ-deception time.

Theorem 5 *For all $D > 0$, and for all attacks in $\mathcal{A}(L)$ using a non-divergent learning algorithm in the exploitation phase, if $\mathbb{E}[T(\epsilon)] \leq D + o(1)$ as $\epsilon \to 0$, and for all $k > L$, the control policy satisfies*

$$\mathbb{E}[V_k^T \hat{A}_k^T \hat{A}_k V_k] + \sigma^2 + 2\mathbb{E}[V_k^T \hat{A}_k^T U_k] \leq 0, \tag{20}$$

then the expected energy of the control signal is

$$R(n_0) \geq \frac{2\sigma^2 \log(1/\epsilon)}{\|\hat{A}_L - A\|_{op}^2 D} + o(\log(1/\epsilon)), \text{ as } \epsilon \to 0, \tag{21}$$

where $R(n_0) \triangleq \mathbb{E}\big[\sum_{k=L}^{n_0-1} U_{k-1}^T U_{k-1}\big]/n_0$.

Theorem 5 shows that the expected energy of the control signal until a time between $L \leq k \leq n_0$ is inversely proportional to the upper bound D on the deception time. Since L is unknown to the controller, it follows that the controller should maintain a high level of expected signal energy $\mathbb{E}[U_k^2]$ at every time instance k to ensure a small deception time.

5. Simulations

In this section, we provide two numerical examples. Although our theoretical results are valid for a large class of learning algorithms and any detection strategy chosen by the controller, we validate them here using LS algorithm and a covariance detector.

First we start with an example for scalar system, where we use the empirical performance of a variance-test to illustrate our results. Specifically, at a decision time τ, the controller tests the empirical variance for unexpected behaviour over a detection window $[0,\tau]$, using a confidence interval $2\gamma > 0$ around the expected variance. More precisely, at decision time τ, $\hat{\Theta}_\tau = 0$ if

$$\frac{1}{\tau}\sum_{k=0}^{\tau}[Y_{k+1} - aY_k - U_k]^2 \in (\mathrm{Var}[W]-\gamma, \mathrm{Var}[W]+\gamma), \tag{22}$$

otherwise $\hat{\Theta}_\tau = 1$. In this case, since the system disturbances are i.i.d. Gaussian $\mathcal{N}(0,\sigma^2)$, using Chebyshev's inequality, we have

$$P_{\mathrm{FA}}^{\tau} \le \frac{\mathrm{Var}[W^2]}{\gamma^2 T} = \frac{3\sigma^4}{\gamma^2 T}. \tag{23}$$

In our simulations, the attacker learns in the exploration phase only, and uses the LS learning algorithm. At the end of the exploration phase, we have

$$\hat{A}_L = \frac{\sum_{k=1}^{L-1}(X_{k+1}-U_k)X_k}{\sum_{k=1}^{L-1}X_k^2}. \tag{24}$$

Our simulation parameters are the following: $\gamma = 0.1$, decision time $\tau = 800$, $A = 1.1$, and $\{W_k\}$ are i.i.d. Gaussian $\mathcal{N}(0,1)$. Using (23), the false-alarm rate is negligible for these parameters.

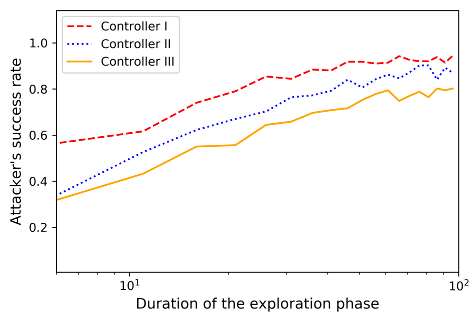
Fig. 2(*a*) compares the attacker's success rate as a function of the duration L of the exploration phase for three different control policies $U_k = -AY_k + \Gamma_k$ such that for all k, I) $\Gamma_k = 0$, II) Γ_k are i.i.d. Gaussian $\mathcal{N}(0,9)$, III) Γ_k are i.i.d. Gaussian $\mathcal{N}(0,16)$. As illustrated in Fig. 2(*a*), the attacker's success rate increases as the duration of exploration phase increases. This is because the attacker's estimation error $|\hat{A}_L - A|$ reduces as L increases, which makes it difficult for the controller to detect the attacker. This is in accordance with the theoretical findings in Theorem 3. Also, for a fixed L, the attacker's success rate decreases as the input control energy increases. The increase in the control energy increases the energy of the fictitious signal which increases the probability of detection, and is in accordance with Theorem 5.

Next, we provide an example of vector system, and analyze the empirical performance of the covariance test against the learning-based attack. In vector systems, the error matrix is

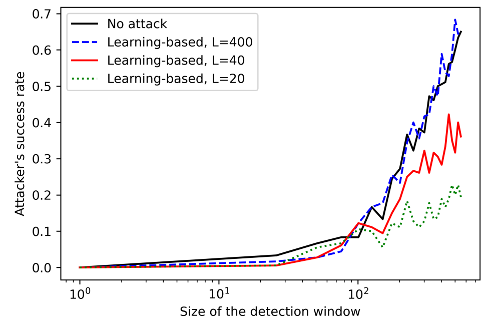
$$\Delta \triangleq \Sigma - \frac{1}{\tau}\sum_{k=1}^{\tau}[Y_{k+1} - AY_k - U_k]\,[Y_{k+1} - AY_k - U_k]^\top$$

Similar to (22), at decision time τ, we have $\hat{\Theta}_\tau = 0$ if $\|\Delta\|_{op} \le \gamma$, and $\hat{\Theta}_\tau = 1$, otherwise. Similar to the scalar system, the attacker learns in the exploration phase only, and uses the LS learning algorithm, which implies that

$$\hat{A}_L = \begin{cases} 0_{n\times n}, & \det(G_{L-1}) = 0; \\ \sum\limits_{k=1}^{L-1}(X_{k+1}-U_k)X_k^\top G_{L-1}^{-1}, & \text{otherwise}, \end{cases} \tag{25}$$



(a) Attacker's success rate versus L



(b) Attacker's success rate versus τ

Figure 2: Simulations Result.

where $G_\tau \triangleq \sum_{k=1}^{\tau} X_k X_k^\top$. Our simulation parameters are the following: $\gamma = 0.1$, $A = [1\ \ 2\,;\ 3\ \ 4]$, $\Sigma = [1\ \ 0\,;\ 0\ \ 1]$, and $U_k = -0.9AY_k$.

Fig. 2(b) compares the attacker's success rate, as a function of sizes of detection window τ for different duration L of the exploration phase. The false-alarm rate decreases to zero as the duration of the τ detection window tends to infinity, similarly to the argument for scalar systems. Thus, as the size of the detection window grows, the success rate of learning-based attacks increases. Finally, as as seen in Fig. 2(b), as the duration of the exploration phase L increases, the attacker's success rate increases, since the attacker improves its estimate of A as L increases. This is in line with the theoretical findings in Theorem 3.

6. Conclusions and Future Directions

We have presented tight lower and upper bounds on the expected deception time for learning based MITM attacks, as the probability of correct detection tends to one. Additionally, we provided an order-optimal characterization of the length of the attacker's exploration phase and computed a lower bound on the control cost. In the future, we plan to study online phase learning based attacks, where the attacker can choose to switch between exploration and exploitation phases dynamically. We also plan to study methods to mitigate these attacks and render the system secure. The extension of our results to partially-observable linear vector systems where the input (actuation) gain is unknown, and the characterization of securable and unsecurable subspaces, similar to Satchidanandan and Kumar (2018), is another possible research direction. Further extensions to nonlinear systems are also of interest.

Acknowledgments

We gratefully acknowledge support from NSF awards CNS-1446891 and ECCS-1917177.

References

Felix Berkenkamp, Matteo Turchetta, Angela Schoellig, and Andreas Krause. Safe model-based reinforcement learning with stability guarantees. In *Advances in neural information processing* *systems*, pages 908–918, 2017.

Ilija Bogunovic, Arpan Losalka, Andreas Krause, and Jonathan Scarlett. Stochastic linear bandits robust to adversarial attacks. In *International Conference on Artificial Intelligence and Statistics*, pages 991–999. PMLR, 2021.

Mona Buisson-Fenet, Friedrich Solowjow, and Sebastian Trimpe. Actively learning gaussian process dynamics. In *Learning for Dynamics and Control*, pages 5–15. PMLR, 2020.

Yiding Chen and Xiaojin Zhu. Optimal attack against autoregressive models by manipulating the environment. *arXiv preprint arXiv:1902.00202*, 2019.

Richard Cheng, Mohammad Javad Khojasteh, Aaron D Ames, and Joel W Burdick. Safe multi-agent interaction through robust control barrier functions with learned uncertainties. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 777–783. IEEE, 2020.

Herman Chernoff. Sequential design of experiments. *The Annals of Mathematical Statistics*, 30(3): 755–770, 1959.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, Aug 2019.

Seyed Mehran Dibaji, Mohammad Pirani, David Bezalel Flamholz, Anuradha M Annaswamy, Karl Henrik Johansson, and Aranya Chakrabortty. A systems and control perspective of cps security. *Annual Reviews in Control*, 47:394–411, 2019.

Joseph L Doob. *Stochastic processes*, volume 7. 1953.

David D Fan, Ali-akbar Agha-mohammadi, and Evangelos A Theodorou. Deep learning tubes for tube mpc. *arXiv preprint arXiv:2002.01587*, 2020.

Aidin Ferdowsi and Walid Saad. Deep learning for signal authentication and security in massive internet-of-things systems. *IEEE Transactions on Communications*, 67(2):1371–1387, 2018.

Jaime F Fisac, Anayo K Akametalu, Melanie N Zeilinger, Shahab Kaynama, Jeremy Gillula, and Claire J Tomlin. A general safety framework for learning-based control in uncertain robotic systems. *IEEE Transactions on Automatic Control*, 64(7):2737–2752, 2018.

David B Flamholz, Anuradha M Annaswamy, and Eugene Lavretsky. Baiting for defense against stealthy attacks on cyber-physical systems. In *AIAA Scitech 2019 Forum*, page 2338, 2019.

Navid Hashemi and Justin Ruths. Gain design via LMIs to minimize the impact of stealthy attacks. In *2020 American Control Conference (ACC)*, pages 1274–1279. IEEE, 2020.

Andreas Hoehn and Ping Zhang. Detection of covert attacks and zero dynamics attacks in cyber-physical systems. In *American Control Conference (ACC), 2016*, pages 302–307. IEEE, 2016.

Yu Jiang and Zhong-Ping Jiang. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. *Automatica*, 48(10):2699–2704, 2012.

Kwang-Sung Jun, Lihong Li, Yuzhe Ma, and Jerry Zhu. Adversarial attacks on stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 3640–3649, 2018.

Aris Kanellopoulos and Kyriakos G Vamvoudakis. A moving target defense control framework for cyber-physical systems. *IEEE Transactions on Automatic Control*, 65(3):1029–1043, 2019.

Mohammad Javad Khojasteh, Anatoly Khina, Massimo Franceschetti, and Tara Javidi. Authentication of cyber-physical systems under learning-based attacks. *IFAC-PapersOnLine*, 52(20): 369–374, 2019.

Mohammad Javad Khojasteh, Vikas Dhiman, Massimo Franceschetti, and Nikolay Atanasov. Probabilistic safety constraints for learned high relative degree system dynamics. In *Learning for Dynamics and Control*, pages 781–792, 2020.

Mohammad Javad Khojasteh, Anatoly Khina, Massimo Franceschetti, and Tara Javidi. Learning-based attacks in cyber-physical systems. *IEEE Transactions on Control of Network Systems*, 8 (1):437–449, 2021.

Armin Lederer, Jonas Umlauft, and Sandra Hirche. Uniform error bounds for gaussian process regression with application to safe control. In *Advances in Neural Information Processing Systems*, pages 659–669, 2019.

Yuzhe Ma, Kwang-Sung Jun, Lihong Li, and Xiaojin Zhu. Data poisoning attacks in contextual bandits. In *International Conference on Decision and Game Theory for Security*, pages 186–204. Springer, 2018.

Yuzhe Ma, Xuezhou Zhang, Wen Sun, and Jerry Zhu. Policy poisoning in batch reinforcement learning and control. In *Advances in Neural Information Processing Systems*, pages 14570–14580, 2019.

Yanbing Mao, Hamidreza Jafarnejadsani, Pan Zhao, Emrah Akyol, and Naira Hovakimyan. Novel stealthy attack and defense strategies for networked control systems. *IEEE Transactions on Automatic Control*, 2020.

Fei Miao, Miroslav Pajic, and George J Pappas. Stochastic game approach for replay attack detection. In *Decision and control (CDC), 2013 IEEE 52nd annual conference on*, pages 1854–1859. IEEE, 2013.

Yilin Mo, Rohan Chabukswar, and Bruno Sinopoli. Detecting integrity attacks on SCADA systems. *IEEE Transactions on Control Systems Technology*, 22(4):1396–1407, 2014.

Yilin Mo, Sean Weerakkody, and Bruno Sinopoli. Physical authentication of control systems: designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Systems*, 35(1):93–109, 2015.

Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, 2013.

Matthew Porter, Pedro Hespanhol, Anil Aswani, Matthew Johnson-Roberson, and Ramanarayan Vasudevan. Detecting generalized replay attacks via time-varying dynamic watermarking. *IEEE Transactions on Automatic Control*, 2020.

Amin Rakhsha, Goran Radanovic, Rati Devidze, Xiaojin Zhu, and Adish Singla. Policy teaching via environment poisoning: Training-time adversarial attacks against reinforcement learning. In *International Conference on Machine Learning*, pages 7974–7984. PMLR, 2020.

Amin Rakhsha, Xuezhou Zhang, Xiaojin Zhu, and Adish Singla. Reward poisoning in reinforcement learning: Attacks against unknown learners in unknown environments. *arXiv preprint arXiv:2102.08492*, 2021.

Anshuka Rangi, Massimo Franceschetti, and Stefano Marano. Consensus-based chernoff test in sensor networks. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 6773–6778. IEEE, 2018a.

Anshuka Rangi, Massimo Franceschetti, and Stefano Marano. Decentralized chernoff test in sensor networks. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 501–505. IEEE, 2018b.

Anshuka Rangi, Massimo Franceschetti, and Stefano Marano. Distributed chernoff test: Optimal decision systems over networks. *arXiv preprint arXiv:1809.04587*, 2018c.

Anshuka Rangi, Massimo Franceschetti, and Stefano Marano. Distributed chernoff test: Optimal decision systems over networks. *IEEE Transactions on Information Theory*, 2020a.

Anshuka Rangi, Mohammad Javad Khojasteh, and Massimo Franceschetti. Learning-based attacks in cyber-physical systems: Exploration, detection, and control cost trade-offs. *arXiv preprint arXiv:2011.10718*, 2020b.

Anshuka Rangi, Long Tran-Thanh, Haifeng Xu, and Massimo Franceschetti. Secure-ucb: Saving stochastic bandits from poisoning attacks via limited data verification. *arXiv preprint arXiv:2102.07711*, 2021.

Anders Rantzer. Concentration bounds for single parameter adaptive control. In *2018 Annual American Control Conference (ACC)*, pages 1862–1866. IEEE, 2018.

Tuhin Sarkar and Alexander Rakhlin. Near optimal finite time identification of arbitrary linear dynamical systems. In *Int. Conf. on Machine Learning (ICML)*, pages 5610–5618, 2019.

Bharadwaj Satchidanandan and Panganamala R Kumar. Dynamic watermarking: Active defense of networked cyber–physical systems. *Proceedings of the IEEE*, 105(2):219–240, 2017.

Bharadwaj Satchidanandan and PR Kumar. Control systems under attack: The securable and unsecurable subspaces of a linear stochastic system. In *Emerging Applications of Control and Systems Theory*, pages 217–228. Springer, 2018.

Yasser Shoukry, Michelle Chong, Masashi Wakaiki, Pierluigi Nuzzo, Alberto Sangiovanni-Vincentelli, Sanjit A Seshia, Joao P Hespanha, and Paulo Tabuada. Smt-based observer design for cyber-physical systems under sensor attacks. *ACM Transactions on Cyber-Physical Systems*, 2(1):5, 2018.

Max Simchowitz, Horia Mania, Stephen Tu, Michael I Jordan, and Benjamin Recht. Learning without mixing: Towards a sharp analysis of linear system identification. *In Conference on Learning Theory*, 2018.

Roy S Smith. Covert misappropriation of networked control systems: Presenting a feedback structure. *IEEE Control Systems*, 35(1):82–92, 2015.

André Teixeira, Iman Shames, Henrik Sandberg, and Karl Henrik Johansson. A secure control framework for resource-limited adversaries. *Automatica*, 51:135–148, 2015.

Sai Vemprala and Ashish Kapoor. Adversarial attacks on optimization based planners. *arXiv preprint arXiv:2011.00095*, 2020.

Draguna Vrabie, O Pastravanu, Murad Abu-Khalaf, and Frank L Lewis. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica*, 45(2):477–484, 2009.

Abraham Wald, Jacob Wolfowitz, et al. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3):326–339, 1948.

Sean Weerakkody, Omur Ozel, Yilin Mo, Bruno Sinopoli, et al. Resilient control in cyber-physical systems: Countering uncertainty, constraints, and adversarial behavior. *Foundations and Trends® in Systems and Control*, 7(1-2):1–252, 2019.

Albert Zhan, Stas Tiomkin, and Pieter Abbeel. Preventing imitation learning with adversarial policy ensembles. *arXiv preprint arXiv:2002.01059*, 2020.

Xuezhou Zhang, Yuzhe Ma, Adish Singla, and Xiaojin Zhu. Adaptive reward-poisoning attacks against reinforcement learning. *arXiv preprint arXiv:2003.12613*, 2020a.

Xuezhou Zhang, Xiaojin Zhu, and Laurent Lessard. Online data poisoning attacks. In *Learning for Dynamics and Control*, pages 201–210, 2020b.

Minghui Zhu and Sonia Martínez. On the performance analysis of resilient networked control systems under replay attacks. *IEEE Transactions on Automatic Control*, 59(3):804–808, 2014.

Xiaojin Zhu. An optimal control view of adversarial machine learning. *arXiv preprint arXiv:1811.04422*, 2018.

Ingvar Ziemann and Henrik Sandberg. Parameter privacy versus control performance: Fisher information regularized control. In *2020 American Control Conference (ACC)*, pages 1259–1265. IEEE, 2020.

Appendix A. Proof of Proposition 1

Proof: Since the attacker does not intervene before $k \leq L$, we have that for all $k \leq L$,

$$D(p_1(Y_{1:k})||p_0(Y_{1:k})) = 0. \tag{26}$$

Thus, for all $k > L$, using the chain rule, we have

$$D(p_1(Y_{1:n})||p_0(Y_{1:n})) = \sum_{k=L+1}^{n} D(p_1(Y_k|Y_{1:k-1})||p_0(Y_k|Y_{1:k-1})). \tag{27}$$

Also, if $\Theta_k = 1$, then for all $k > L$, we have

$$Y_k|(Y_{k-1}, U_{k-1}, \hat{A}_{k-1}) \sim \mathcal{N}(\hat{A}_{k-1}Y_{k-1} + U_{k-1}, \sigma^2 I_M), \tag{28}$$

since $Y_k = V_k$ for all $k > L$. Similarly, if $\Theta_k = 0$, then for all $k > L$, we have

$$Y_k|(Y_{k-1}, U_{k-1}, \hat{A}_{k-1}) \sim \mathcal{N}(AY_{k-1} + U_{k-1}, \sigma^2 I_M). \tag{29}$$

The result now follows by using the fact that for all $k > L$, we have $Y_k = V_k$.

We continue by noticing that the control input U_k lies in sigma field of past observations, namely U_k is measurable with respect to sigma field generated by $Y_{1:k-1}$. Thus, combining (27), (28) and (29), for all $k > L$, we have that

$$D(p_1(Y_k|Y_{1:k-1})||p_0(Y_k|Y_{1:k-1})) = \mathbb{E}\left[\frac{Y_{k-1}^T(\hat{A}_{k-1} - A)^T(\hat{A}_{k-1} - A)Y_{k-1}}{2\sigma^2}\right]. \tag{30}$$

Using (27) and (30), for all $n > L$, we have

$$D(p_1(Y_{1:n})||p_0(Y_{1:n})) = \mathbb{E}\left[\sum_{k=L+1}^{n} \frac{Y_{k-1}^T(\hat{A}_{k-1} - A)^T(\hat{A}_{k-1} - A)Y_{k-1}}{2\sigma^2}\right]. \tag{31}$$

□

Appendix B. Proof of the Theorem 1

Proof: The proof of the theorem consists of two parts. First, for all attacks in the class $\mathcal{A}(L)$ and $0 < c < 1$, we show that if the probability of error is small, namely $\mathbb{P}(\hat{\Theta}_\tau \neq \Theta_\tau) = O(|\epsilon \log \epsilon|)$, then the log-likelihood ratio $\log(p_1(y_{1:\tau})/p_0(y_{1:\tau}))$ should be greater than $(1-c)\log(1/\epsilon)$ with high probability as $\epsilon \to 0$, namely

$$\log \frac{p_1(y_{1:\tau})}{p_0(y_{1:\tau})} \geq (1-c)\log(1/\epsilon) \tag{32}$$

must hold with high probability, as $\epsilon \to 0$. Second, we show that there exists $0 < \bar{c} < 1$ such that for all $0 < c \leq \bar{c}$ and $T(\epsilon) < (1-c)\log(1/\epsilon)/C(n_0)$, it is unlikely that the inequality in (32) is satisfied.

Using (10), for all $k \geq L+1$, we have that both type I and type II errors of the hypothesis test $\Theta_k = 1$ vs. $\Theta_k = 0$ are $O(|\epsilon \log \epsilon|)$. Thus, using (Chernoff, 1959, Lemma 4), for all $0 < c < 1$, we have

$$\mathbb{P}\Big(S^\tau \leq -(1-c)\log \epsilon\Big) = O(-\epsilon^c \log \epsilon), \tag{33}$$

where

$$S^n = \log \frac{p_1(y_{1:n})}{p_0(y_{1:n})} = \sum_{k=1}^{n} \log \left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})} \right). \tag{34}$$

Therefore, as $\epsilon \to 0$, the probability in (33) tends to 0, which concludes the first part of the proof.

Now, we show that for all $0 < c < 1$, we have

$$\lim_{n' \to \infty} \mathbb{P}\left(\max_{1 \leq k \leq n'} S^k \geq (D(p_1(y_{1:n'})||p_0(y_{1:n'})) + n'c) \right) = 0, \tag{35}$$

where $D(p_1(y_{1:n'})||p_0(y_{1:n'}))$ denotes the KL divergence between the distributions p_1 and p_0 of $Y_{1:n'}$. We have

$$\begin{aligned} S^n &= \sum_{k=1}^{n} \left(\log \left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})} \right) - D(p_1(Y_k|Y_{1:k-1})||p_0(Y_k|Y_{1:k-1})) \right) \\ &\quad + \sum_{k=1}^{n} D(p_1(Y_k|Y_{1:k-1})||p_0(Y_k|Y_{1:k-1})) \\ &= M_1^n + M_2^n, \end{aligned} \tag{36}$$

where

$$M_1^n = \sum_{k=1}^{n} \left(\log \left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})} \right) - D(p_1(Y_k|Y_{1:k-1})||p_0(Y_k|Y_{1:k-1})) \right), \tag{37}$$

is a martingale with mean 0 with respect to filtration $\mathcal{F}_k = \sigma(Y_{1:k-1})$, and

$$\begin{aligned} M_2^n &= \sum_{k=1}^{n} D(p_1(Y_k|Y_{1:k-1})||p_0(Y_k|Y_{1:k-1})), \\ &\overset{(a)}{=} D(p_1(Y_{1:n})||p_0(Y_{1:n})), \end{aligned} \tag{38}$$

where (a) follows from the chain rule of KL-Divergence. Now, if the event in (35) occurs for a fixed n_1, namely

$$M_1^{n_1} + M_2^{n_1} \geq D(p_1(Y_{1:n_1})||p_0(Y_{1:n_1})) + n_1 c, \tag{39}$$

then it implies that $M_1^{n_1} \geq n_1 c$. Since $Y_k|Y_{1:k-1}$ has a normal distribution using (28) and (29), there exists a constant $b > 0$ such that the probability in (35) simplifies as

$$\mathbb{P}\left(\max_{1 \leq k \leq n'} S^k \geq (D(p_1(y_{1:n'})||p_0(y_{1:n'})) + n'c) \right) \leq \mathbb{P}(\max_{1 \leq k \leq n'} M_1^k \geq n'c) \overset{(a)}{\leq} b/n'c^2, \tag{40}$$

where (a) follows from the Doob-Kolmogorov extension of Chebyshev's inequality in Doob (1953), and the fact that M_1^k is a martingale with 0 mean. Hence, we have that (35) follows.

Now, we have

$$n_0 C(n_0) < \log(1/\epsilon). \tag{41}$$

Therefore, there exists $0 < \bar{c} < 1$ such that

$$n_0 C(n_0) + n_0 \bar{c} = (1 - \bar{c}) \log(1/\epsilon). \tag{42}$$

Now, using Proposition 1, for all $0 < c \leq \bar{c}$, we have

$$\begin{aligned}\mathbb{P}(N \leq n_0) &\leq \mathbb{P}\Big(N \leq n_0 \text{ and } S^N \geq n_0(C(n_0) + c)\Big) + \mathbb{P}\Big(S^N \leq n_0(C(n_0) + c)\Big)\\ &\leq \mathbb{P}\Big(\max_{1\leq k\leq n_0} S^k \geq n_0(C(n_0) + c)\Big) + \mathbb{P}\Big(S^N \leq n_0(C(n_0) + c)\Big),\end{aligned} \tag{43}$$

and the first and the second terms at the right-hand side of (43) approach zero by (35) and (33), respectively. □

Appendix C. Proof of the Theorem 2

Proof: In $\mathcal{A}(L)$, consider an attack R^* such that for all $k > L$, we have $\hat{A}_k = \hat{A}_L$. For all $k > L$, if $\Theta_k = 1$, then we have

$$Y_k | Y_{1:k-1} \sim \mathcal{N}(\hat{A}_L Y_{k-1} + U_{k-1}, \sigma^2 I_M). \tag{44}$$

Similarly, if $\Theta_k = 0$, then

$$Y_k | Y_{1:k-1} \sim \mathcal{N}(A Y_{k-1} + U_{k-1}, \sigma^2 I_M). \tag{45}$$

Consider a the following detection strategy, also known as Sequential Probability Ratio Test (SPRT), at the controller as follows. At time n, if

$$\sum_{k=1}^{n} \log\left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})}\right) \geq \log(1/\epsilon), \tag{46}$$

then $\hat{\Theta}_n = 1$, and if

$$\sum_{k=1}^{n} \log\left(\frac{p_0(y_k|y_{1:k-1})}{p_1(y_k|y_{1:k-1})}\right) \geq \log(1/\epsilon), \tag{47}$$

then $\hat{\Theta}_n = 0$. Otherwise, n is not a decision time and the test continues.

We will show that for the attack R^* and the detection strategy SPRT, the statement of the theorem holds.

For SPRT, the probability of error, both P^{τ}_{Dec} and P^{τ}_{FA}, is at most $O(\epsilon)$, and the proof is along the same direction as (Rangi et al., 2018c, Theorem 1). Now, let us prove the bound on $T(\epsilon)$. Given the system is under attack, let the decision time τ of SPRT be

$$T = \min\left\{n : \sum_{k=1}^{n} \log\left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})}\right) \geq \log(1/\epsilon)\right\}. \tag{48}$$

Using (Chernoff, 1959, Lemma 2), for system under attack $\mathcal{A}(L)$ and for all $c > 0$, there exist a $b > 0$ such that

$$\mathbb{P}\left(\sum_{k=1}^{n} \log\left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})}\right) < (D(p_1(Y_{1:n})||p_0(Y_{1:n})) - nc)\right) \leq e^{-bn}. \tag{49}$$

Using the definition of n_0, for all $\bar{n} > n_0$ we have

$$\log(1/\epsilon) \leq \bar{n}C(\bar{n}) = D(p_1(Y_{1:\bar{n}})||p_0(Y_{1:\bar{n}})), \tag{50}$$

where the equality follows from Proposition 1. Using (49) and (50), For all $c > 0$ and $n \geq (1 + c)(n_0 + 1)\log(1/\epsilon)/D(p_1(Y_{1:n_0+1})||p_0(Y_{1:n_0+1}))$, we have

$$\mathbb{P}\left(\sum_{k=1}^{n} \log\left(\frac{p_1(y_k|y_{1:k-1})}{p_0(y_k|y_{1:k-1})}\right) < \log(1/\epsilon)\right) \leq e^{-bn}. \tag{51}$$

Then, using Proposition 1, the statement of the theorem follows. □

Appendix D. Proof of Theorem 3

Proof: If the learning algorithm in the exploration phase is a convergent algorithm, the learning algorithm in the exploitation phase is a non-divergent algorithm, then for all $0 < \delta < 1$, we have

$$\begin{aligned} C(n_0) &\overset{(a)}{\leq} \|\hat{A}_L - A\|_{op}^2 \frac{1}{n_0} \mathbb{E}\left[\sum_{k=L+1}^{n_0} \frac{V_{k-1}^T V_{k-1}}{2\sigma^2}\right], \\ &\overset{(b)}{\leq} \left(\frac{c^{1/\alpha}}{L\delta^{1/\alpha}}\right)\tilde{C}(n_0), \end{aligned} \tag{52}$$

with probability at least $1 - \delta$, where (a) follows from the fact that

$$||Ax||_2 \leq ||A||_{op}||x||_2, \tag{53}$$

and the learning algorithm in the exploitation phase is non-divergent, and (b) follows from Definition 6 of convergent algorithms. Thus, we have

$$\frac{\log(1/\epsilon)}{C(n_0)} \geq \frac{\log(1/\epsilon)}{\tilde{C}(n_0)}\left(\frac{L\delta^{1/\alpha}}{c^{1/\alpha}}\right), \tag{54}$$

with probability at least $1 - \delta$. Using Theorem 1 and (54), if

$$(1 + o(1))\frac{\log(1/\epsilon)}{\tilde{C}(n_0)}\left(\frac{L\delta^{1/\alpha}}{c^{1/\alpha}}\right) > D(1 + o(1)), \text{ as } \epsilon \to 0, \tag{55}$$

then $\mathbb{E}[T(\epsilon)] > D + o(1)$ as $\epsilon \to 0$. This along with (54) implies that

$$L \geq \frac{(1 + o(1))D\tilde{C}(n_0)}{\log(1/\epsilon)}\frac{c^{1/\alpha}}{\delta^{1/\alpha}}, \text{ as } \epsilon \to 0, \tag{56}$$

with probability at least $1 - \delta$. □

Appendix E. Proof of Theorem 4

Proof: Consider the LS learning algorithm in Simchowitz et al. (2018) which satisfies

$$\mathbb{P}(\|\hat{A}_k - A\|_{op} > \eta) \leq \frac{c_1}{e^{\eta^2 k}}, \tag{57}$$

For $\eta = \sqrt{\log(c_1/\delta)/L}$, similar to (52), we have that

$$C(n_0) \leq \frac{\log(c_1/\delta)}{L}\tilde{C}(n_0), \tag{58}$$

with probability at least $1-\delta$. Thus, we have

$$\frac{\log(1/\epsilon)}{C(n_0)} \geq \frac{\log(1/\epsilon)}{\tilde{C}(n_0)} \frac{L}{\log(c_1/\delta)}, \tag{59}$$

with probability at least $1-\delta$. Thus, for $L = (1+o(1))D\tilde{C}(n_0)\log(c_1/\delta)/\log(1/\epsilon)$ as $\epsilon \to 0$, using Theorem 1, we have that

$$\mathbb{E}[T(\epsilon)] \geq \frac{(1+o(1))\log(1/\epsilon)}{C(n_0)} \geq D(1+o(1)) = D + o(1), \text{ as } \epsilon \to 0, \tag{60}$$

with probability at least $1-\delta$. The statement of the theorem follows. $\square$

Appendix F. Proof of Theorem 5

Proof: Since $\tilde{W}_k$ is independent of U_k and V_k and $\mathbb{E}[\tilde{W}_k] = 0$, we have

$$\mathbb{E}[V_{k+1}^T V_{k+1}] - \mathbb{E}[U_k^T U_k] = \mathbb{E}[V_k^T \hat{A}_k^T \hat{A}_k V_k] + \sigma^2 + 2\mathbb{E}[V_k^T \hat{A}_k^T U_k]. \tag{61}$$

Using (20), we have

$$\mathbb{E}[V_{k+1}^T V_{k+1}] \leq \mathbb{E}[U_k^T U_k], \tag{62}$$

which implies

$$\begin{aligned} C(n_0) &\overset{(a)}{\leq} \frac{\|\hat{A}_L - A\|_{op}^2}{n_0} \mathbb{E}\left[\sum_{k=L+1}^{n_0} \frac{V_{k-1}^T V_{k-1}}{2\sigma^2}\right] \\ &\overset{(b)}{\leq} \frac{\|\hat{A}_L - A\|_{op}^2}{n_0} \mathbb{E}\left[\sum_{k=L}^{n_0-1} \frac{U_{k-1}^T U_{k-1}}{2\sigma^2}\right], \end{aligned} \tag{63}$$

where (a) follows from the fact that $||Ax||_2 \leq ||A||_{op}||x||_2$, and (b) follows from (62). Since $\mathbb{E}[T(\epsilon)] \leq D + o(1)$ as $\epsilon \to 0$, using Theorem 1 and (63), we have that

$$D + o(1) \geq \frac{(1+o(1))2\sigma^2 \log(1/\epsilon)}{\|\hat{A}_L - A\|_{op}^2 R(n_0)}, \text{ as } \epsilon \to 0. \tag{64}$$

Hence, the statement of the theorem follows. $\square$