Automatica 82 (2017) 251–260

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs☆

Cheng-Zong Bai [a], Fabio Pasqualetti [b], Vijay Gupta [a]

[a] Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN, USA
[b] Department of Mechanical Engineering, University of California, Riverside, CA, USA

ARTICLE INFO

Article history:
Received 12 April 2016
Received in revised form
3 April 2017
Accepted 17 April 2017

Keywords:
Cyber–physical system security
Networked control systems
Stochastic systems

ABSTRACT

Consider a stochastic process being controlled across a communication channel. The control signal that is transmitted across the control channel can be replaced by a malicious attacker. The controller is allowed to implement any arbitrary detection algorithm to detect if an attacker is present. This work characterizes some fundamental limitations of when such an attack can be detected, and quantifies the performance degradation that an attacker that seeks to be undetected or stealthy can introduce.

1. Introduction

Using communication channels to inject malicious data that degrades the performance of a cyber–physical system has now been demonstrated both theoretically and practically (Farwell & Rohozinski, 2011; Kuvshinkova, 2003; Mo, Chabukswar, & Sinopoli, 2014; Pasqualetti, Dörfler, & Bullo, 2013; Richards, 2008; Slay & Miller, 2007). Intuitively, there is a tradeoff between the performance degradation an attacker can induce and how easy it is to detect the attack (Teixeira, Pérez, Sandberg, & Johansson, 2012). Quantifying this tradeoff is of great interest to operate and design secure cyber–physical systems (CPS).

As explained in more detail later, for noiseless systems, zero dynamics provide a fundamental notion of stealthiness of an attacker, which characterizes the ability of an attacker to stay undetected even if the controller can perform arbitrary tests on the data it receives. However, similar notions for stochastic systems have been lacking. In this work, we consider stochastic cyber–physical systems, propose a graded stealthiness notion, and characterize the performance degradation that an attacker with a given level of stealthiness can induce. The proposed notion is fundamental in the sense that we do not constraint the detection test that the controller can employ to detect the presence of an attack.

Related work Security of cyber–physical systems is a growing research area. Classic works in this area focus on the detection of sensor and actuator failures in control systems (Patton, Frank, & Clark, 1989), whereas more recent approaches consider the possibility of intentional attacks at different system layers; e.g., see Pasqualetti, Dörfler, and Bullo (2015). Both simple attacks, such as jamming of communication channels (Foroush & Martínez, 2013), and more sophisticated attacks, such as replay and data injection attacks, have been considered (Mo & Sinopoli, 2010; Smith, 2011).

One way to organize the literature in this area is based on the properties of the considered cyber–physical systems. While initial studies focused on static systems (Dan & Sandberg, 2010; Giani et al., 2011; Liu, Reiter, & Ning, 2009; Mohsenian-Rad & Leon-Garcia, 2011; Teixeira, Amin, Sandberg, Johansson, & Sastry, 2010), later works exploited the dynamics of the system either to design attacks or to improve the performance of the detector that a controller can employ to detect if an attack is present (Bhattacharya & Başar, 2013; Hamza, Tabuada, & Diggavi, 2011; Maharjan, Zhu, Zhang, Gjessing, & Başar, 2013; Manshaei, Zhu, Alpcan, Başar, & Hubaux, 2011; Zhu & Martínez, 2011; Zhu, Tembine, & Başar, 2013). For noiseless cyber–physical systems, the concept of stealthiness of an attack is closely related to the control-theoretic notion of zero dynamics (Basile & Marro, 1991, Section 4).

☆ Work supported in part by awards NSF ECCS-1405330 and ONR N00014-14-1-0816. A preliminary version of this work appeared in Bai et al. (2015). With respect to Bai et al. (2015), this paper (i) considers more general systems with multiple inputs and outputs, (ii) completes and extends technical proofs, and (iii) provides further insight into the design of optimal stealthy attacks in stochastic cyber–physical systems. The material in this paper was partially presented at the 2015 American Control Conference, July 1–3, 2015, Chicago, IL, USA. This paper was recommended for publication in revised form by Associate Editor Hideaki Ishii under the direction of Editor Christos G. Cassandras.

E-mail addresses: cbai@nd.edu (C.-Z. Bai), fabiopas@engr.ucr.edu (F. Pasqualetti), vgupta2@nd.edu (V. Gupta).

http://dx.doi.org/10.1016/j.automatica.2017.04.047
0005-1098/

In particular, an attack is undetectable in noiseless systems if and only if it excites only the zero dynamics of an appropriately defined input–output system describing the system dynamics, the measurements available to the security monitor, and the variables compromised by the attacker (Fawzi, Tabuada, & Diggavi, 2014; Pasqualetti et al., 2013). For cyber–physical systems driven by noise, instead, the presence of process and measurements noise offers the attacker an additional possibility to tamper with sensor measurements and control inputs within acceptable uncertainty levels, thereby making the detection task more difficult.

Detectability of attacks in stochastic systems remains an open problem. Most works in this area consider detectability of attacks with respect to specific detection schemes employed by the controller, such as the classic bad data detection algorithm (Cui et al., 2012; Mo & Sinopoli, 2010). The trade-off between stealthiness and performance degradation induced by an attacker has also been characterized only for specific systems and detection mechanisms (Kosut, Jia, Thomas, & Tong, 2011; Kwon, Liu, & Hwang, 2013; Liu, Ning, & Reiter, 2011; Mo et al., 2014), and a thorough analysis of resilience of stochastic control systems to arbitrary attacks is still missing. While convenient for analysis, the restriction to a specific class of detectors prevents the characterization of fundamental detection limitations. In our previous work (Bai & Gupta, 2014), we proposed the notion of ϵ-marginal stealthiness to quantify the stealthiness level in an estimation problem with respect to the class of ergodic detectors. In this work, we remove the assumption of ergodicity and introduce a notion of stealthiness for stochastic control systems that is independent of the attack detection algorithm, and thus provides a fundamental measure of the stealthiness of attacks in stochastic control systems. Further, we also characterize the performance degradation that such a stealthy attack can induce.

We limit our analysis to linear, time-invariant plants with a controller based on the output of an asymptotic Kalman filter, and to injection attacks against the actuation channel only. Our choice of using controllers based on Kalman filters is not restrictive. In fact, while this is typically the case in practice, our results and analysis are valid for arbitrary control schemes. Our choice of focusing on attacks against the actuation channel only, instead, is motivated by two main reasons. First, actuation and measurements channels are equally likely to be compromised, especially in networked control systems where communication between sensors, actuators, plant, and controller takes place over wireless channels. Second, this case has received considerably less attention in the literature – perhaps due to its enhanced difficulty – where most works focus on attacks against the measurement channel only; e.g., see Fawzi et al. (2014) and Teixeira et al. (2010). We remark also that our framework can be extended to the case of attacks against the measurement channel, as we show in Bai and Gupta (2014) for scalar systems and a different notion of stealthiness.

Finally, we remark that since the submission of this work, some recent literature has appeared that builds on it and uses a notion of attack detectability that is similar to what we propose in Bai and Gupta (2014), Bai, Pasqualetti, and Gupta (2015) and in this paper. For instance, Kung, Dey, and Shi (2016) extend the notion of ϵ-stealthiness of Bai et al. (2015) to higher order systems, and show how the performance of the attacker may differ in the scalar and vector cases (in this paper we further extend the setup in Kung et al. (2016) by leveraging the notion of right-invertibility of a system to consider input and output matrices of arbitrary dimensions). In Zhang and Venkitasubramaniam (2016), the authors extend the setup in Bai et al. (2015) to vector and not necessarily stationary systems, but consider a finite horizon problem. In Guo, Shi, Johansson, and Shi (2017), the degradation of remote state estimation is studied, for the case of an attacker that compromises the system measurements based on a linear strategy. Two other relevant recent works are Weerakkody, Sinopoli, Kar, and Datta (2016) that use the notion of Kullback-Liebler divergence as a causal measure of information flow to quantify the effect of attacks on the system output, while Chen, Kar, and Moura (2016) characterize optimal attack strategies with respect to a linear quadratic cost that combines attackers control and undetectability goals.

Contributions The main contributions of this paper are threefold. First, we propose a notion of ϵ-stealthiness to quantify detectability of attacks in stochastic cyber–physical systems. Our metric is motivated by the Chernoff–Stein lemma in detection and information theories and is universal because it is independent of any specific detection mechanism employed by the controller. Second, we provide an information theoretic bound for the degradation of the minimum-mean-square estimation error caused by an ϵ-stealthy attack as a function of the system parameters, noise statistics, and information available to the attacker. Third, we characterize optimal stealthy attacks, which achieve the maximal degradation of the estimation error covariance for a stealthy attack. For right-invertible systems (Basile & Marro, 1991, Section 4.3.2), we provide a closed-form expression of optimal ϵ-stealthy attacks. The case of single-input single-output systems considered in our conference paper (Bai et al., 2015) is a special case of this analysis. For systems that are not right-invertible, we propose a sub-optimal ϵ-stealthy attack with an analytical expression for the induced degradation of the system performance. We include a numerical study showing the effectiveness of our bounds. Our results provide a quantitative analysis of the trade-off between performance degradation that an attacker can induce versus a fundamental limit of the detectability of the attack.

Paper organization Section 2 contains the mathematical formulation of the problems considered in this paper. In Section 3, we propose a metric to quantify the stealthiness level of an attacker, and we characterize how this metric relates to the information theoretic notion of Kullback–Leibler Divergence. Section 4 contains the main results of this paper, including a characterization of the largest performance degradation caused by an ϵ-stealthy attack, a closed-form expression of optimal ϵ-stealthy attacks for right invertible systems, and a suboptimal class of attacks for not right-invertible systems. Section 5 presents illustrative examples and numerical results. Finally, Section 6 concludes the paper.

2. Problem formulation

Notation: The sequence $\{x_n\}_{n=i}^{j}$ is denoted by x_i^j (when clear from the context, the notation x_i^j may also denote the corresponding vector obtained by stacking the appropriate entries in the sequence). This notation allows us to denote the probability density function of a stochastic sequence x_i^j $f_{x_i^j}$, and to define its differential entropy $h(x_i^j)$ as (Cover & Thomas, 2006, Section 8.1)

$$h(x_i^j) \triangleq \int_{-\infty}^{\infty} -f_{x_i^j}(t_i^j) \log f_{x_i^j}(t_i^j) dt_i^j.$$

Let x_1^k and y_1^k be two random sequences with probability density functions (pdf) $f_{x_1^k}$ and $f_{y_1^k}$, respectively. The Kullback–Leibler Divergence (KLD) (Cover & Thomas, 2006, Section 8.5) between x_1^k and y_1^k is defined as

$$D(x_1^k \parallel y_1^k) \triangleq \int_{-\infty}^{\infty} \log \frac{f_{x_1^k}(t_1^k)}{f_{y_1^k}(t_1^k)} f_{x_1^k}(t_1^k) dt_1^k. \tag{1}$$

The KLD is a non-negative quantity that gauges the dissimilarity between two probability density functions with $D(x_1^k \parallel y_1^k) = 0$