Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 52-20 (2019) 387–390

Optimal sensor design for secure cyber-physical systems

Mohamed Ali Belabbas[1] **, Xudong Chen**[2] *

* *Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, IL, USA and Electrical and Computer Engineering Department, CU Boulder, CO, USA*

Abstract:
The vision of secure cyber-physical systems requires to develop a new array of tools to prevent, monitor for, and recover from a wide variety of potential attacks to a system. We contribute to this vision here by addressing the problem of optimal sensor design for detection of injection attacks. More precisely, we consider an attack in which a sensor output is corrupted, and we have at our disposal a sensor deployed in an ad-hoc fashion and thus is secure. How to design the sensor so that we maximize the probability of detection of such an injection attack? We pose the problem here as an hypothesis test, describe an optimization problem that the optimal sensor needs to satisfy, and furthermore obtain the analytic solution in a particular case.

Keywords: Sensor design, Secure sensing, Neyman-Pearson test, Cyber-physical systems

1. INTRODUCTION

We consider here the optimal placement of sensors to detect intrusive signals in control systems. In general, one design sensors to maximize the quality of the estimate of a signal of interest, and work done in Belabbas [2016], Chen [2018] showed that this problem is tractable in some regimes. Here, we consider situations in which one suspects that a sensor may be breached and subject to signal injection. How could we detect such a sensor attack? The type of scenarios we have in mind is the following: consider a plant described by a linear dynamics with additive Gaussian noise. The state x of the plant is observed through p linear sensors $C_i x$, subject to sensor noise, and from the output of these sensors, a Kalman-Bucy filter estimate $\hat{x}$ of the state is obtained. One has access to a secured sensor, this sensor can be built in the plant or can be set-up in an ad-hoc fashion when one suspects an attack is taking place for example. The general question we ask is then the following: how to best design the sensors of the plant and the secure sensor to maximize the probability of detection of a signal injection, while keeping the false alarm rate (deciding there is an attack when no attack is taking place) under a given value?

Common control theoretic approaches to security are often related to notions of *robustness* using tools from robust control theory Bamieh et al. [1999], Dullerud and Paganini [2000], Doyle et al. [2013], or that of *fault tolerance* and the study of fault detection and isolation in dynamical systems. In this context, significant attention has been placed on distributed averaging protocols (consensus) corrupted by noises Xiao et al. [2007], Das et al. [2010], Hatano et al. [2005], Zelazo and Mesbahi [2011]. These works provided analysis results relating the performance of these systems to the underlying network structure using an $\mathcal{H}_2$ performance measure Xiao et al. [2007], Zelazo and Mesbahi [2011], Bamieh et al. [2012], Patterson and Bamieh [2011].

For studies with a focus on sensor design, we mention Sayin and Başar [2017] where secure sensor design under channel tampering is studied from a game theoretic point of view. In Fawzi et al. [2011], the authors investigate the problem of state estimations in CPS when sensors are under attack. For other game-theoretic approaches, we refer to Li et al. [2017] and Saad et al. [2011], where cooperation among wireless nodes has been recently proposed for improving the physical layer security of wireless transmission in the presence of multiple eavesdroppers. Taking the point of view of the attacker, the authors of Kim et al. [2015] use data driven, subspace methods to design attacks on state estimators.

2. PROBLEM FORMULATION

We now give a general problem statement, of which only a particular case will be solved below. Consider the control system with observation:

$$\begin{cases} dx = Axdt + Budt + dw, \\ dy_0 = C_0 xdt + dv_0, \\ dy_i = C_i xdt + dv_i + s_i dt, \quad \forall i = 1, \dots, p, \end{cases} \tag{1}$$

where $y_0(t)$ is the measurement output from the secured sensor and other $y_i(t)$'s, for all $i = 1, \dots, p$, are the measurement outputs from the sensors of the plant. Throughout the paper, we assume that A is stable, so that a pair (A, C) is always detectable Brockett [2015]. The signals s_i are injection attacks (and s_i can be 0). The problem we are concerned with in the paper is two-fold: (1) How

[1] belabbas@illinois.edu, [2] xudong.chen@colorado.edu. M.-A. Belabbas would like to acknowledge the support of grant NSF ECCS-1809076. X. Chen would like to acknowledge the support of grant NSF ECCS-1809315.

2405-8963 © 2019, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2019.12.186

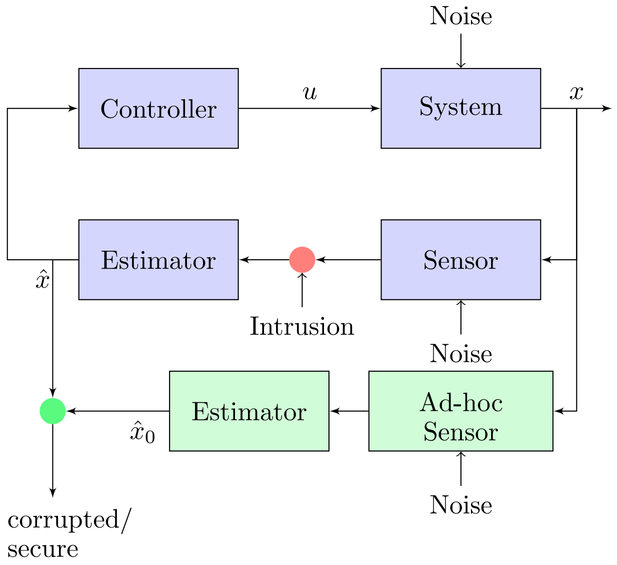

Fig. 1. If an attack is suspected, an ad-hoc sensor (in green) or secure sensor can be used to form an estimate $\hat{x}_0$ of the state that is compare to the estimate $\hat{x}$. We derive the sensor and decision rule that yield the highest detection rate of attacks.

can we utilize the secured sensor to detect injection attacks? We will propose an approach below which compares the Kalman-Bucy filter estimates of the states using the secured sensor and the sensors of the plant. Of course, always declaring that the system is under attack will yield a perfect detection rate; however, it is likely to also result in a very high false alarm rate. Hence, one needs to set a bound on a maximum allowable false alarm rate. This results in a standard Neyman-Pearson test; (2) Based on the proposed approach, how can we design all the sensors $C_0, \ldots, C_p$ (of fixed norm) to maximize the detection rate of injection attacks under a given false alarm rate.

The rationale behind (1) is that given sensor signals y_i, most algorithms will run a Kalman-Bucy filter to obtain an estimate $\hat{x}$ to be used for monitoring or control purposes, and thus this estimate will be readily available in most practical situations. The rational behind (2) is the realization that not all sensors are the same when it comes to detecting an injection attack, and thus given the freedom to design a sensor, one should seek the one that maximizes the detection rate for a given false alarm rate.

2.1 Detection approach

We now elaborate on technical details of the above mentioned detection approach. For ease of presentation, we will consider the special case where $p = 1$, i.e., there is only one sensor of plant. As is argued above, we will compare two Kalman-Bucy filter estimates $\hat{x}_0(t)$ and $\hat{x}_1(t)$, which use $y_0(t)$ and $y_1(t)$ respectively. More specifically, since A is stable, both (A, C_0) and (A, C_1) are detectable. Then, it is well known that $\hat{x}_0(t)$ and $\hat{x}_1(t)$, for t sufficiently large (so that the asymptotic regime is considered), are given by the following:

$$\begin{cases} d\hat{x}_i = A\hat{x}_i dt + \Sigma_i C_i^\top (dy_i - C_i \hat{x}_i dt) \\ 0 = A\Sigma_i + \Sigma_i A^\top + I - \Sigma_i C_i^\top C_i \Sigma_i \end{cases} \tag{2}$$

where we assume, without loss of generality, that $B = 0$ in Eq. (1). Now, let

$$z(t) := \hat{x}_1(t) - \hat{x}_0(t). \tag{3}$$

Note that $z(t)$ is a Gaussian random variable, whose mean and variance can be computed as follows:

Absence of injection attack: In this case, we have that $s_1(t) \equiv 0$ and it should be clear that the mean of $z(t)$ is 0 in steady-state:

$$\mathbb{E}z(t) \equiv 0. \tag{4}$$

The covariance of $z(t)$ (in steady-state as well) is given by

$$\operatorname{cov}(z(t)) = \Sigma_0 + \Sigma_1. \tag{5}$$

To see this, we note that

$$\begin{aligned} z(t) &= (\hat{x}_1(t) - x(t)) - (\hat{x}_0(t) - x(t)) \\ &:= e_1(t) - e_0(t), \end{aligned}$$

where $e_i(t)$ is the estimation error of each Kalman-Bucy filter estimate and it obeys the following differential equation:

$$de_i = (A - \Sigma_i C_i^\top C_i)e_i + \Sigma_i C_i^\top dv_i, \quad \forall i = 0, 1, \tag{6}$$

From (6), we see that $e_0(t)$ and $e_1(t)$ are independent random variables; hence, the covariance of $z(t)$ is the sum of the covariances of $e_0(t)$ and $e_1(t)$.

Presence of injection attack: The injection attack can be either deterministic or stochastic, but we assume in the latter case that $s_1(t)$ is independent of $v_i(t)$. Compared to the previous case, the effect of $s_1(t)$ on the $z(t)$ is to shift the mean of $z(t)$ to

$$\mathbb{E}z(t) = \Sigma_1 C_1^\top \mathbb{E}s_1(t), \tag{7}$$

and the covariance of $z(t)$ to

$$\operatorname{cov}(z(t)) = \Sigma_0 + \Sigma_1 + \operatorname{var}(s_1(t))\Sigma_1 C_1^\top C_1 \Sigma_1.$$

Of course, when $s_1(t)$ is deterministic, then $\operatorname{cov}(s_1(t)) = 0$ and, hence, $\operatorname{cov}(z(t))$ will be reduced to $\Sigma_1 + \Sigma_0$ which is the same as the covariance in the previous case.

The general case can be hard to tackle. We focus in the paper on a simple case where the injection attack

$$s_1(t) \equiv s \tag{8}$$

is constant. Such an assumption significantly simplifies the detection problem. On the other hand, the simplification helps illustrate the optimal sensor design problem, which is the main focus of the paper.

To this end, we now recall the Neyman-Pearson test:

Background: Neyman-Pearson test: We denote by H_0 the hypothesis that there is no injection attack on the sensor, and by H_1 the hypothesis that there is an injection attack. An attack detector is then a binary function $\mathbb{I}(z(t))$ which takes the value 1 when we believe that H_1 is true, and 0 otherwise. The *detection rate* D of $\mathbb{I}$ is then

$$D := \mathbb{P}(\mathbb{I}(z(t)) = 1 \mid H_1 \text{ is true}),$$

and the false alarm rate F is

$$F := \mathbb{P}(\mathbb{I}(z(t)) = 1 \mid H_0 \text{ is true}).$$

Set $0 < \alpha < 1$ to be the maximal allowable false detection rate; we seek to find Λ so that D is maximized with $F \leq \alpha$. The Neyman-Pearson lemma Neyman and Pearson [1933] states that the above optimum is obtained by comparing the ratio of likelihood of both hypotheses with a threshold determined by the false alarm rate. To be more specific, let λ_i be the likelihood that H_i is true given $z(t)$:

$$\lambda_i := \mathbb{P}(H_i \text{ is true} \mid z(t)).$$

Now set $\Lambda(z)$ to be the ratio of these likelihoods:

$$\Lambda(z) := \frac{\lambda_0}{\lambda_1}.$$

The Neyman-Pearson lemma states that there exists a threshold η so that setting

$$\mathbb{I}(z(t)) = \begin{cases} 0 \text{ if } \Lambda(z(t)) \geq \eta, \\ 1 \text{ otherwise.} \end{cases}$$

yields the optimal detector. Furthermore, η is obtained by solving the equation

$$\alpha = \mathbb{P}(\Lambda(z) \leq \eta \mid H_0 \text{ is true}).$$

Sensor design for attack detection: We now formulate in precise terms what the optimal sensor design for attack detection is. Given the false alarm rate F, the detection rate D will depend on the the sensors C_0 and C_1. We maximize the detection rate through an optimal design of the two sensors. We can obtain the following theorem, whose proof we omit here due to space constraints:

Theorem 2.1. Consider the Neyman-Pearson test described above with given false alarm rate F. Set

$$\Phi : (C_0, C_1) \mapsto C_1 \Sigma_1 (\Sigma_0 + \Sigma_1)^{-1} \Sigma_1 C_1^\top.$$

Then the sensors C_0^*, C_1^* of unit norm that maximize the detection rate D are

$$(C_0^*, C_1^*) = \arg \max_{\|C_i\|=1} \Phi(C_0, C_1),$$

where Σ_i satisfy the ARE in Eq. (2).

2.2 A conjecture and an analytic solution

The above theorem provides a strong characterization of the optimal sensors, using which, numerical simulations can be performed. The numerical solutions led us to the following conjecture:

Conjecture 2.1. The optimal sensor placement for injection attack detection is so that $C_1^* = C_0^*$. More precisely,

$$\arg \max_{\|C_i\|=1} \Phi(C_0, C_1) = \arg \max_{\|C\|=1} \Phi(C, C).$$

The conjecture is more or less natural: in order to decide whether the sensor C_1 has been affected by an injection attack, it is better to compare it to a similar sensor C_0 which is likely to yield a measurement output $y_0(t)$ that is statistically similar to $y_1(t)$. However, this is not apparent from the expression of the optimal sensor given in Theorem 2.1 since there, the roles of C_0 and C_1 are not symmetric, i.e. $\Phi(C_0, C_1)$ is not necessarily equal to $\Phi(C_1, C_0)$.

We will assume in the sequel that Conjecture 2.1 is true. For ease of notation, we let $C := C_0 = C_1$ and since $C_0 = C_1$, it follows that $\Sigma_0 = \Sigma_1 =: \Sigma$. Consequently, the expression of Φ can be simplified as follows:

$$\Phi(C) := C \Sigma C^\top / 2.$$

We note again that Σ depends on C through the algebraic Riccati equation.

Our goal now is to maximize the above $\Phi(C)$ over $\|C\| = 1$. The optimization problem is still hard to solve. A geometric approach to tackle the problem is to derive the gradient flow of the cost function Φ over the space of rank-one matrices $\{CC^\top \mid \|C\| = 1\}$ and investigate the equilibrium points of the gradient flow. These equilibrium points necessarily include the global minimum point. Of course, a significant amount of efforts will be made to analyze these equilibrium points and decided which one could be a local/global minimum point. More details of such an approach can be found in Belabbas [2016] and Chen [2018].

We take here a different approach, which relies mostly on matrix analysis. However, at the stage, we cannot handle the most general case. We can only provide a complete solution to the special case where A is symmetric. We summarize below the main result:

Theorem 2.2. Let $A \in \mathbb{R}^{n \times n}$ be stable and symmetric. Let $0 > \lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of A and $v_1, \ldots, v_n$ be the corresponding eigenvectors of unit length. Then,

$$\arg \max_{\|C\|=1} \Phi(C) = v_1^\top,$$

and

$$\max_{\|C\|=1} \Phi = \frac{1}{2(\sqrt{\lambda_1^2 + 1} - \lambda_1)}.$$

We note that Theorem 2.2 can be straightforwardly generalized to the case where $\|C\| = r$ is arbitrary: The optimal C is aligned with v_1, and $\max_{\|C\|=r} \Phi$ becomes

$$\max_{\|C\|=r} \Phi = \frac{r^2}{2(\sqrt{\lambda_1^2 + r^2} - \lambda_1)}. \tag{9}$$

3. CONCLUSION AND OUTLOOK

Cyber-physical systems are subject, by their very nature, to a wide variety of potential attacks, and it is not always easy of possible to prevent or predict such attacks. In these cases, monitoring the system to detect anomalous behaviors is of the utmost importance, since early detection can help avoid catastrophic failure. Hence, finding methods that provide the best detection rate of attacks is thus similarly of the utmost importance. We addressed this problem in this paper in the case of an injection attack on a sensor, and provided an expression for optimal sensors (optimal in the sense that they maximize the detection rate given a bound on the false alarm rate). Furthermore, we gave an explicit solution to the problem in a particular case, showing that the optimal placement problem reduced to an eigen-problem, and provided a conjecture as to the general case.

The analysis performed here is however still limited to the case of a unique sensor, and it would be of course quite important to extend it to the case of several sensors. Furthermore, the optimization problem here only addresses injection attack detection, ignoring the quality of the sensor for estimation of the state of the system, its primary purpose. Obtaining sensors which balance these two objectives, namely quality of estimation and attack detection rate, is also an important problem which we believe the techniques outlined here would allow us to solve.

REFERENCES

Bamieh, B., Jovanovic, M.R., Mitra, P., and Patterson, S. (2012). Coherence in Large-Scale Networks: Dimension-Dependent Limitations of Local Feedback. *IEEE Transactions on Automatic Control*, 57(9), 2235–2249.

Bamieh, B., Paganini, F., and Dahleh, M.A. (1999). *Optimal control of distributed arrays with spatial invariance*, volume 245 of *Lecture Notes in Control and Information Sciences*, chapter 24, 329–343. Springer London, London.

Belabbas, M.A. (2016). Geometric methods for optimal sensor design. *Proceedings of the Royal Society, Series A Math Phys Eng Sci*, 472(2185), 20150312.

Brockett, R.W. (2015). *Finite dimensional linear systems*, volume 74. SIAM.

Chen, X. (2018). Joint actuator-sensor design for stochastic linear systems. In *2018 IEEE Conference on Decision and Control (CDC)*, 6668–6673. IEEE.

Das, A., Hatano, Y., and Mesbahi, M. (2010). Agreement over noisy networks. *IET Control Theory & Applications*, 4(11), 2416.

Doyle, J.C., Francis, B.A., and Tannenbaum, A.R. (2013). *Feedback control theory*. Courier Corporation.

Dullerud, G. and Paganini, F. (2000). *A course in robust control theory: a convex approach*. Springer-Verlag, New York.

Fawzi, H., Tabuada, P., and Diggavi, S. (2011). Secure state-estimation for dynamical systems under active adversaries. In *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, 337–344. IEEE.

Hatano, Y., Mesbahi, M., and Das, A. (2005). Agreement in presence of noise: pseudogradients on random geometric networks. In *44th IEEE Conference on Decision and Control, 2005*, 6382–6387. Seville, Spain.

Kim, J., Tong, L., and Thomas, R.J. (2015). Subspace methods for data attack on state estimation: A data driven approach. *IEEE Trans. Signal Processing*, 63(5), 1102–1114.

Li, Y., Quevedo, D.E., Dey, S., and Shi, L. (2017). A game-theoretic approach to fake-acknowledgment attack on cyber-physical systems. *IEEE Transactions on Signal and Information Processing over Networks*, 3(1), 1–11.

Neyman, J. and Pearson, E.S. (1933). Ix. on the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706), 289–337.

Patterson, S. and Bamieh, B. (2011). Network Coherence in Fractal Graphs. In *Proc. 50th IEEE Conference on Decision and Control*, 6445–6450. Orlando, FL.

Saad, W., Han, Z., Başar, T., Debbah, M., and Hjørungnes, A. (2011). Distributed coalition formation games for secure wireless transmission. *Mobile Networks and Applications*, 16(2), 231–245.

Sayin, M.O. and Başar, T. (2017). Secure sensor design for cyber-physical systems against advanced persistent threats. In *International Conference on Decision and Game Theory for Security*, 91–111. Springer.

Xiao, L., Boyd, S., and Kim, S. (2007). Distributed average consensus with least-mean-square deviation. *Journal of Parallel and Distributed Computing*, 67, 33–46. doi: 10.1016/j.jpdc.2006.08.010.

Zelazo, D. and Mesbahi, M. (2011). Edge Agreement: Graph-Theoretic Performance Bounds and Passivity Analysis. *IEEE Transactions on Automatic Control*, 56(3), 544–555.