2014 American Control Conference (ACC)
June 4-6, 2014. Portland, Oregon, USA

Finite Energy and Bounded Attacks on Control System Sensor Signals

Seddik M. Djouadi, Alexander M. Melin, Erik M. Ferragut, Jason A. Laska, and Jin Dong

Abstract— **Control system networks are increasingly being connected to enterprise level networks. These connections leave critical industrial controls systems vulnerable to cyber-attacks. Most of the effort in protecting these cyber-physical systems (CPS) from attacks has been in securing the networks using information security techniques. Effort has also been applied to increasing the protection and reliability of the control system against random hardware and software failures. However, the inability of information security techniques to protect against all intrusions means that the control system must be resilient to various signal attacks for which new analysis methods need to be developed.**

In this paper, sensor signal attacks are analyzed for observer-based controlled systems. The threat surface for sensor signal attacks is subdivided into denial of service, finite energy, and bounded attacks. In particular, the error signals between states of attack free systems and systems subject to these attacks are quantified. Optimal sensor and actuator signal attacks for the finite and infinite horizon linear quadratic (LQ) control in terms of maximizing the corresponding cost functions are computed. The closed-loop systems under optimal signal attacks are provided. Finally, an illustrative numerical example using a power generation network is provided together with distributed LQ controllers.

I. INTRODUCTION

When a control application failure can cause irreparable harm to people or the physical system being controlled, it is called safety-critical [1]. In particular, supervisory control and data acquisition (SCADA) systems perform vital functions in electric power production and distribution, oil and natural gas production and distribution, water and wastewater treatment, transportation systems, health-care devices, chemical production, and weapon systems to name a few [1]. Attacks on these systems can have significant impact on public safety and cause colossal economic losses.

SCADA and cyber-physical (CP) networks are increasingly connected to enterprise networks and are beginning to utilize wireless sensor and actuator networks to improve functionality and reduce cost. For example, the smart grid program [2] uses extensive wireless technology. All these factors increase the risk and several specific intentional attacks on control systems have been reported in the literature, e.g., [3], [5]–[7], [10]. Most of the effort for protecting control systems has been in protection against random failures [1], [13]. However, there is a pressing need in protecting these systems against malicious attacks [14]–[16].

Critical control systems have relied on physical barriers and network airgaps for protection [17], [18]. The Stuxnet virus has shown that these protections are inadequate as it was able to bypass both the physical security and a network airgap with relative ease [10]. Anonymity is also insufficient protection as more high profile viruses attacking control systems are found [11], [12], [17].

Numerous studies and real-world incidents have shown that the existing security methods are ineffective to guarantee safe and reliable operations of CPSs against unexpected attacks [19], [20]. Information security methods, such as authentication and cryptography methods, appear inadequate for protecting CPSs [19], since these security methods do not exploit the compatibility of the measurements with the underlying physical process and controller [21], [22] and can affect the performance and stability of the control system. In [24] deception and denial of service attacks against a networked control system are introduced, and a countermeasure based on semidefinite programming is proposed. In [25] false data injection attacks against static state estimators are shown possible with limited resources. Similarly, stealthy deception attacks against SCADA systems are studied in [24], [26]. In [27] the effect of replay attacks on a control system is analyzed. In [28] the effect of covert attacks against networked control systems is investigated. In [29] a resilient control problem where an attacker corrupts control packets is discussed and receding-horizon control law is proposed to stabilize the control system. In [22] common information of the physical system state is exploited for the key establishment between the sensor and the controller. In [30] a trustiness system is introduced to the controller, which computes the trustiness of different sensors.

Rigorous definitions of state awareness, operational normalcy, and resiliency of control systems are proposed in [17], [31]. Robust and resilient control techniques applied to CP systems have been reported in [32], [33].

In this paper, sensor signal attacks are analyzed for observer-based controlled systems. In particular, the error

S. Djouadi is an Associate Professor with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996, djouadi@eecs.utk.edu

A. Melin is with the Sensors and Controls Research Group, Oak Ridge National Laboratory, Oak Ridge, TN 37831, melina@ornl.gov

E. Ferragut is with the Cyberspace Sciences and Information Intelligence Research Group, Oak Ridge National Laboratory, Oak Ridge, TN 37831, ferragutem@ornl.gov

J. Laska is with the Cyberspace Sciences and Information Intelligence Research Group, Oak Ridge National Laboratory, Oak Ridge, TN 37831, laskaja@ornl.gov

J. Dong is with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996, jdong@utk.edu

Research sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. The submitted manuscript has been authored by a contractor of the U.S. Government under Contract DE-AC05-00OR22725. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

978-1-4799-3274-0/$31.00 ©2014 AACC

signals between states of attack free systems and systems subject to these attacks are quantified. Optimal sensor signal attacks for the finite and infinite horizon linear quadratic (LQ) control in terms of maximizing the corresponding cost functions are computed. The closed-loop systems under optimal signal attacks are provided.

II. Sensor Cyber-Attacks

In this section, cyber-attacks on the sensor measurements are analyzed. We assume that the attacker is capable of spoofing sensor signals with a time-varying signal $\Delta_y(t)$ that starts at $t = 0$. We assume that (A, B) is controllable, and (A, C) is observable so there exists a matrix L such that the eigenvalues of $(A - LC)$ are in the open left half plane, i.e., $eig(A - LC) < 0$. The system and controller based observer have the following form:

$$\begin{array}{rcl} \dot{x}(t) & = & Ax(t) + Bu(t) \\ y_\alpha(t) & = & Cx(t) + \Delta_y(t) \\ \dot{\hat{x}}(t) & = & A\hat{x}(t) + Bu(t) + L(y_\alpha(t) - C\hat{x}(t)) \\ u(t) & = & -K\hat{x}(t) + Gr(t) \end{array} \qquad (1)$$

where $r(t)$ is the reference signal and the matrix G is chosen to remove steady state tracking error. Define the error signal $\hat{e}(t) := x(t) - \hat{x}(t)$, then the closed-loop system has the following state-space representation:

$$\begin{array}{l} \begin{pmatrix} \dot{\hat{x}}(t) \\ \dot{\hat{e}}(t) \end{pmatrix} = \begin{pmatrix} A - BK & LC \\ 0 & A - LC \end{pmatrix} \begin{pmatrix} \hat{x}(t) \\ \hat{e}(t) \end{pmatrix} \\ \qquad + \begin{pmatrix} BG \\ 0 \end{pmatrix} r(t) + \begin{pmatrix} L \\ -L \end{pmatrix} \Delta_y(t) \end{array} \qquad (2)$$

The case where the bias signal Δ_y is constant has been studied in [17], [18]. Note that if the attacker has complete access to the measurements and enough resources to cancel them out, i.e., $\Delta_y(t) = -Cx(t), \forall t$, then it is straightforward to show that such an attack can be destabilizing if the original system is unstable. A simple example is

$$\dot{x}(t) = x(t) + u(t) \qquad (3)$$

$$y(t) = 2x(t) \qquad (4)$$

Here $A = 1,\ B = 1,\ C = 2$. This system is unstable. To stabilize it choose, for e.g., $K = 2$, $L = 1$, so $A - BK = -1$, $A - LC = -1$. By choosing $\Delta_y = -2x(t)$, the spoofed sensor signal becomes $y_\alpha(t) = 0$ and the system has a pole at 1 and therefore is unstable. This is representative of a *denial of service* attack on the control system although other DoS attacks are possible such as packet erasure or corruption.

A. Finite Energy Attacks

Another scenario is to assume that the attacker can inject a bounded or finite energy signal in time, $\Delta_y \in L^2([0, \infty))$, where $L^2([0, \infty))$ the space of Lebesgue measurable and square integrable functions, i.e.,

$$\|\Delta_y\|_2^2 := \int_0^\infty \|\Delta_y(t)\|^2 dt < \infty \qquad (5)$$

where $\| \cdot \|$ stands for the Euclidean norm. Let us see how the attack signal $\Delta_y(\cdot)$ affect the error signal $\hat{e}(\cdot)$. Solving (2) for $\hat{e}(\cdot)$ yields:

$$\hat{e}(t) = \mathrm{e}^{(A-LC)t}\hat{e}(0) + \int_0^t -\mathrm{e}^{(A-LC)(t-\tau)} L\Delta_y(\tau) d\tau \qquad (6)$$

Taking the Laplace transform yields:

$$\begin{array}{l} E(s) = (s - A + LC)^{-1}\hat{e}(0) \\ \qquad - (s - A + LC)^{-1} L\Delta_Y(s) \end{array} \qquad (7)$$

where $E(\cdot)$ is the Laplace transform of $\hat{e}(t)$. Since $\Delta_y \in L^2[0, \infty)$, its Laplace transform $\Delta_Y(s) \in H^2$, where H^2 is the Hardy space of analytic and square integrable functions in the open right half-plane. To determine $\hat{e}(t)$ in steady-state it suffices to apply the final value theorem:

$$\begin{array}{rl} \lim\limits_{t \longrightarrow \infty} \hat{e}(t) & = \lim\limits_{s \longrightarrow 0} sE(s) \\ & = \lim\limits_{s \longrightarrow 0} s\Big((s - A + LC)^{-1}\hat{e}(0) \\ & \quad - (s - A + LC)^{-1} L\Delta_Y(s)\Big) = 0 \end{array} \qquad (8)$$

Expression (8) shows that finite energy sensor attacks are mitigated in steady state by the controller based observer scheme.

B. Bounded Attack Signals

The third attack scenario is a bounded attack signal. For bounded attack signals, i.e. $\Delta_y \in L^\infty[0, \infty)$, $\operatorname{ess\,sup}_{t \in [0,\infty)} \|\Delta_y(t)\| \leq \delta$, for some $\delta \geq 0$ the Laplace transform is not necessarily defined. We have for all $t > 0$:

$$\|\hat{e}(t)\| \leq \|\mathrm{e}^{(A-LC)t}\hat{e}(0)\| + \delta \int_0^t \|\mathrm{e}^{(A-LC)(t-\tau)} L\| d\tau, \qquad (9)$$

Since the eigenvalues of $A - LC$ have negative real parts, $\|(e)^{(A-LC)t}\| \leq c\ \mathrm{e}^{\lambda t},\ \forall t \geq 0,\ \exists\ c \geq 0,\ \lambda < 0$, and in steady state we have:

$$\lim_{t \longrightarrow \infty} \|\hat{e}(t)\| \leq -\frac{c\delta}{\lambda}\ \|L\| \qquad (10)$$

Expression (10) shows that under an L^∞-signal attack the error $\hat{e}(t)$ is bounded by a bias proportional to the attack signal bound, however the error can be persistent and result in deterioration in system performance, in that, the estimated and actual states stay apart. In the next, section, optimal attacks for finite horizon linear quadratic control are derived.

III. Optimal Sensor Attacks on Finite Horizon Linear Quadratic (LQ) Control

We consider the plant described by the following state-space system:

$$\begin{array}{l} \dot{x}(t) = Ax(t) + B_2 u(t), \quad x(0) = x_0 \\ z(t) = \begin{pmatrix} C_1\, x \\ u \end{pmatrix} \end{array} \qquad (11)$$

The finite horizon linear quadratic (LQ) problem is concerned with minimizing the cost function:

$$J(u, x_0, h, Q) = \int_0^h z^T(\tau) z(\tau) d\tau + x^T(h) Q x(h) \qquad (12)$$

where Q is a positive semi-definite matrix, $Q \geq 0$. The objective of the LQ controller is the minimization of the cost (12) over causal linear full-information controllers. From standard LQ theory the optimal controller is the state feedback [34]:

$$u = -B_2^T P x \tag{13}$$

where P is the solution of the Riccati equation:

$$-\dot{P} = PA + A^T P - P B_2 B_2^T P + C_1^T C_1, \quad P(h) = Q \tag{14}$$

The matrix P is non negative semi-definite, $P \geq 0$ and is bounded above for any $\tau \leq h$ [34]. By completing the square the cost $J(u, x_0, h, Q)$ takes the form:

$$\begin{aligned} J(u, x_0, h, Q) &= x_0^T P(0) x_0 \\ &+ \int_0^h (u + B_2^T P x)^T (u + B_2^T P x) d\tau \end{aligned} \tag{15}$$

where, with no attack, the optimal controller is given by (13), and the optimal cost:

$$J^\star(u, x_0, h, Q) = x_0^T P(0) x_0 \tag{16}$$

In the next section, the effect of sensor signal attacks on LQ control system is studied.

A. Optimal Sensor Attack

Since we are assuming full-information the measurement signal $y(t) = x(t)$, and a sensor attack takes the form:

$$y_\alpha(t) = x(t) + \Delta_y(t), \quad t \geq 0 \tag{17}$$

In the presence of the sensor attack signal $\Delta_y(t)$ we write the cost function as $J^\star(u, x_0, h, Q, \Delta_y)$. From (15) such an attack increases the cost to:

$$\begin{aligned} J^\star(u,&x_0, h, Q, \Delta_y) = \\ &= \big(x_0 + \Delta_y(0)\big)^T P(0) \big(x_0 + \Delta_y(0)\big) \\ &+ \int_0^h \|B_2^T P \Delta_y\|^2 d\tau \end{aligned} \tag{18}$$

Thus any sensor attack satisfies:

$$\begin{aligned} J^\star(u,&x_0, h, Q, \Delta_y) \leq \\ &\lambda_{\max}\big(P(0)\big) \big(x_0 + \Delta_y(0)\big)^T \big(x_0 + \Delta_y(0)\big) \\ &+ \int_0^h \|B_2^T P(\tau)\|^2 \|\Delta_y(\tau)\|^2 d\tau \end{aligned} \tag{19}$$

where $\lambda_{\max}\big(P(0)\big)$ is the maximal eigenvalue of $P(0)$, the first term on the RHS follows from the Rayleigh-Ritz inequality. The maximal singular value of the matrix $B_2^T P(\tau)$ at τ is $\|B_2^T P(\tau)\|^2 \equiv \lambda_{\max}\big(P(\tau) B_2 B_2^T P(\tau)\big)$. Therefore, from the performance point of view the worst case sensor attack corresponds to the eigenvector, denoted $v_1(0) \in \mathbb{C}^n$, associated to $\lambda_{\max}\big(P(0)\big)$, and the (normalized) singular vector, denoted $v_1(\tau) \in \mathbb{C}^n$, $\tau > 0$, associated to $\|B_2^T P(\tau)\|$, that is:

$$\begin{aligned} P(0) v_1(0) &= \lambda_{\max}\big(P(0)\big) v_1(0), \\ P(\tau) B_2 B_2^T P(\tau) v_1(\tau) &= \|B_2^T P(\tau)\|^2 v_1(\tau), \\ \|v_1(0)\| &= 1, \\ \|v_1(\tau)\| &= 1, \ \tau > 0 \end{aligned} \tag{20}$$

Identities (20) give the direction of the worst case sensor attack signals as $\Delta_y(0) = \alpha_0 v_1(0) - x(0)$, for some scalar α_0, and $\Delta_y(\tau) = \alpha(\tau) v_1(\tau)$ for some scalar-valued function $\alpha(\cdot)$. These can be easily determined by an attacker if the system parameters are known and will increase the cost to:

$$\begin{aligned} \sup_{\Delta_y} \ & J^\star(u, x_0, h, Q, \Delta_y) \\ &= \sup_{\alpha_0} \alpha_0^2 \lambda_{\max}\big(P(0)\big) \\ &+ \sup_{\alpha(\cdot)} \int_0^h \alpha^2(\tau) \|B_2^T P(\tau)\|^2 d\tau \end{aligned} \tag{21}$$

Constrains must be placed on α_0 and $\alpha(\cdot)$ otherwise the supremum on the RHS of (21) would be infinite and realistically, an unbounded attack signal is not feasible. Potential attacks include special classes of signals such as finite energy, i.e. L^2-signals, and bounded L^∞-signals.

B. Finite Energy Optimal Sensor Attacks Over A Finite Time Horizon

Let us first consider the former by assuming that $\Delta_y(\cdot) \in L^2([0,h), \mathbb{R}^n)$, where $L^2([0,h), \mathbb{R}^n)$ is the space of finite energy functions defined in the interval $[0,h)$ and taking values in $\mathbb{R}^n$. We shall assume that the sensor attack cannot expand more than an energy threshold, say $|\alpha_0| \leq M$ and $\|\Delta_y\|_2 \leq M$. Since $P(\cdot)$ is bounded in $[0,h)$, $B_2^T P(\cdot)$ can be viewed as a multiplication operator from $L^2([0,h), \mathbb{R}^n)$ into $L^2([0,h), \mathbb{R}^m)$. Then the supremum of the second term in (18) can be computed as the operator induced norm:

$$\begin{aligned} \sup_{\|\Delta_y\|_2 \leq M} & \int_0^h \|B_2^T P \Delta_y\|^2 d\tau \\ &= \sup_{\tau \in [0,h)} M^2 \|B_2^T P(\tau)\|^2 \end{aligned} \tag{22}$$

and the worst case cost is:

$$\begin{aligned} \sup_{\Delta_y} \ & J^\star(u, x_0, h, Q, \Delta_y) \\ &= M^2 \lambda_{\max}\big(P(0)\big) \\ &+ M^2 \sup_{\tau \in [0,h]} \lambda_{\max}\big(P(\tau) B_2 B_2^T P(\tau)\big) \end{aligned} \tag{23}$$

To compute a sensor attack Δ_y that results in (23), let $B_2^T P(\tau)$ have singular values arranged in decreasing order $\sigma_1(\tau) \geq \sigma_2(\tau) \geq \cdots \geq \sigma_m(\tau)$, $\tau \in [0,h]$, and singular value decomposition $B_2^T P(\tau) = U(\tau) D(\tau) V^\star(\tau)$, $\forall \tau \in [0,h]$, where U and V are unitary matrix-valued functions, $V^\star$ is the conjugate transpose (i.e., adjoint) of V, and $D(\tau)$ is the diagonal matrix:

$$D(\tau) = \operatorname{diag}\big(\sigma_1(\tau), \sigma_2(\tau), \cdots, \sigma_m(\tau)\big), \ \tau \in [0,h] \tag{24}$$

By the definition of supremum, for $\epsilon > 0$, there exists a subset E_ϵ such that for each $\tau_\epsilon \in E_\epsilon$, $\|B_2^T P(\tau_\epsilon)\| > \sup_{\tau \in [0,h)} \|B_2^T P(\tau)\| - \epsilon$. Let $\mu(E_\epsilon) > 0$ be a positive Lebesgue measure on the subset E_ϵ and let χ_{E_ϵ} be the characteristic function of E_ϵ, i.e.,

$$\chi_{E_\epsilon}(\tau) = \begin{cases} 1 & \tau \in E_\epsilon \\ 0 & \tau \notin E_\epsilon \end{cases} \tag{25}$$

and define $\xi_\epsilon(\tau) = \frac{1}{\sqrt{\mu(E_\epsilon)}}(\chi_{E_\epsilon}, 0, \cdots, 0)^T$. Then $\xi_\epsilon \in L^2([0,h), \mathbb{R}^n)$, and

$$\|\xi_\epsilon\|_2^2 = \frac{1}{\mu(E_\epsilon)} \int_{E_\epsilon} d\mu = \frac{\mu(E_\epsilon)}{\mu(E_\epsilon)} = 1. \tag{26}$$

Therefore, taking the sensor attack signal:

$$\Delta_y(\tau) = MV(\tau)\xi_\epsilon(\tau),\ \tau \in [0,h) \tag{27}$$

yields:

$$\begin{aligned} &\|B_2^T P(\tau)\Delta_y(\tau)\|_2^2 \\ &\quad = M^2 \int_0^h \|U(\tau)D(\tau)V^\star(\tau)V(\tau)\xi_\epsilon(\tau)\|^2 d\tau \end{aligned} \tag{28}$$

Since $U(\tau)$ and $V(\tau)$ are unitary for each τ, we get:

$$\begin{aligned} \|B_2^T P(\tau)\Delta_y(\tau)\|_2^2 &= M^2 \frac{1}{\mu(E_\epsilon)} \int_{E_\epsilon} \sigma_1^2(\tau) d\tau \\ &> M^2 \frac{1}{\mu(E_\epsilon)} \int_{E_\epsilon} \big(\sup_{\tau \in [0,h]} \|B_2^T P(\tau)\|^2 - \epsilon \big) d\tau \\ &> M^2 \big(\sup_{\tau \in [0,h]} \|B_2^T P(\tau)\|^2 - \epsilon \big) \end{aligned} \tag{29}$$

Since ϵ is arbitrary we have:

$$\|B_2^T P(\tau)\Delta_y(\tau)\|_2^2 \geq M^2 \sup_{\tau \in [0,h)} \|B_2^T P(\tau)\|^2 \tag{30}$$

Expression (30) shows that the sensor signal attack is

$$\begin{aligned} \Delta_y(\tau) &= MV(\tau)\xi_\epsilon(\tau) \\ &= MV(\tau) \frac{1}{\sqrt{\mu(E_\epsilon)}} (\chi_{E_\epsilon}, 0, \cdots, 0)^T,\ \tau \in (0,h) \\ &= M \frac{\sigma_1(\tau)\chi_{E_\epsilon}}{\sqrt{\mu(E_\epsilon)}} v_1(\tau),\ \tau \in (0,h) \end{aligned} \tag{31}$$

where $v_1(\cdot)$ is the first column of $V(\cdot)$ and represents the right singular vector of $B_2^T P(\cdot)$ associated to $\sigma_1(\cdot)$. The sensor signal attack (31) results in the worst deterioration in the LQ cost function. The smaller ϵ the poorer the LQ performance. From the attacker's perspective the finite energy sensor attack (31) is optimal. The corresponding LQ cost function is:

$$\begin{aligned} &\sup_{\|\Delta_y\|_2 \leq M} J^\star(u, x_0, h, Q, \Delta_y) = \\ &\qquad M^2 \lambda_{\max}\big(P(0)\big) + M^2 \|B_2^T P\|_\infty^2 \end{aligned} \tag{32}$$

The corresponding closed-loop system under attack becomes:

$$\begin{aligned} \dot{x}(t) &= (A - B_2 B_2^T P(t))x(t) - B_2 B_2^T P(t)\Delta_y(t),\quad t \geq 0 \\ &= (A - B_2 B_2^T P(t))x(t) \\ &\quad - \frac{M}{\sqrt{\mu(E_\epsilon)}} B_2 U(t)D(t)V^\star(t)V(t)(\chi_{E_\epsilon}(t), 0, \cdots, 0)^T \\ &= (A - B_2 B_2^T P(t))x(t) - \frac{M\sigma_1(t)\chi_{E_\epsilon}(t)}{\sqrt{\mu(E_\epsilon)}} B_2 u_1(t) \end{aligned} \tag{33}$$

where $u_1(t)$ is the first column of $U(t)$ and represents the unit left singular vector corresponding to the maximal singular value $\sigma_1(t)$. Integrating (33) yields:

$$\begin{aligned} x(t) &= M\Phi(t,0)v_1(0) \\ &\quad + \frac{M}{\sqrt{\mu(E_\epsilon)}} \int_0^t \chi_{E_\epsilon}(\tau)\sigma_1(\tau)\Phi(t,\tau)B_2 u_1(\tau) d\tau,\ t \in (0,h) \\ &\geq M\Phi(t,0)v_1(0) \\ &\quad + \frac{M}{\sqrt{\mu(E_\epsilon)}} \Big(\operatorname{ess\,sup}_{t \in [0,h)} \big(\sigma_1(t)\big) - \epsilon \Big) \\ &\quad \times \int_{E_\epsilon \cap [0,t)} \Phi(t,\tau) B_2 u_1(\tau) d\tau \end{aligned} \tag{34}$$

where $\Phi(t,\tau)$ is the state transition matrix associated to the matrix $A - B_2 B_2^T P(t)$. The closed-loop state-space solution reveals three properties:

1) The optimal finite energy sensor attack at time t occurs on a set $E_\epsilon \cap [0,t)$ where the norm if $B_2^T P(t)$, $\sigma_1(t)$, is within ϵ of its maximal magnitude on $[0,h)$.
2) There is a trade-off between the magnitude of the attack and its duration. The smaller ϵ is, the closer $\sigma_1(t)$ is to its peak $\operatorname{ess\,sup}_{t \in [0,h)}\big(\sigma_1(t)\big)$, i.e., the bigger the attack amplitude, but the smaller its duration set E_ϵ since by definition:

$$\begin{aligned} E_\epsilon = \{ & t \in [0,h) : \\ & \sigma_1(t) > \operatorname{ess\,sup}_{t \in [0,h)} \big(\sigma_1(t)\big) - \epsilon,\ \epsilon > 0 \Big\} \end{aligned} \tag{35}$$

   and therefore if $0 < \epsilon_1 < \epsilon_2$ then necessary $E_{\epsilon_1} \subset E_{\epsilon_2}$. Hence the assertion.
3) The attack occurs only on a subset E_ϵ of the horizon $[0,h)$ due to the presence of χ_{E_ϵ}.

C. Bounded Optimal Sensor Attacks Over A Finite Time Horizon

For bounded signal sensor attacks $\Delta_y \in L^\infty([0,h), \mathbb{R}^n)$, under the norm $\|\Delta_y\|_\infty := \operatorname{ess\,sup}_{\tau \in [0,h)} \|\Delta_y(\tau)\|$, the optimal attack strategy is to attempt to maximize the cost function $J^\star(\cdot)$ by maximizing the second term in (18), i.e.,

$$\sup_{\|\Delta_y\|_\infty \leq M} \int_0^h \|B_2^T P \Delta_y\|^2 d\tau \tag{36}$$

Now, we have the following inequalities for any sensor attack signal $\Delta_y \in L^\infty([0,h), \mathbb{R}^n)$, with $\|\Delta_y\|_\infty \leq M$:

$$\begin{aligned} \int_0^h \|B_2^T P \Delta_y\|^2 d\tau &\leq \int_0^h \|B_2^T P(\tau)\|^2 \|\Delta_y(\tau)\|^2 d\tau \\ &\leq M^2 \int_0^h \|B_2^T P(\tau)\|^2 d\tau \end{aligned} \tag{37}$$

The upper bound can be achieved by choosing the attack signal Δ_y as:

$$\begin{aligned} \Delta_y(\tau) &= MV(\tau)(1, 0, \cdots, 0)^T \\ &= Mv_1(\tau) \Longrightarrow \|\Delta_y\|_\infty = M, \end{aligned} \tag{38}$$

since $V(\tau)$ is unitary for all $\tau \in [0,h)$. Indeed, substituting Δ_y in (37) yields:

$$\int_0^h \|B_2^T P(\tau)\Delta_y\|^2 d\tau = M^2 \int_0^h \sigma_1^2(\tau)d\tau = M^2 \int_0^h \|B_2^T P(\tau)\|^2 d\tau \tag{39}$$

showing that the bounded sensor attack signal (38) results in the worst LQ cost function,

$$\sup_{\|\Delta_y\|_\infty \leq M} J^\star(u, x_0, h, Q, \Delta_y) = M^2 \lambda_{\max}\big(P(0)\big) + M^2 \int_0^h \|B_2^T P(\tau)\|^2 d\tau \tag{40}$$

and is optimal from the attacker's standpoint. Similar computations to (34) show that under (38) the closed-loop state-space system under the optimal bounded sensor attack becomes:

$$\dot{x}(t) = (A - B_2 B_2^T P(t))x(t) - M\sigma_1(t)B_2 u_1(t), \tag{41}$$

for $0 \leq t < h$. Integrating yields

$$x(t) = M\Phi(t,0)v_1(0) - M\int_0^t \sigma_1(\tau)\Phi(t,\tau)B_2 u_1(\tau)d\tau \tag{42}$$

for $0 < t < h$.

IV. Optimal Sensor Attacks for Infinite Horizon LQ Control

In the infinite horizon case we let $h \longrightarrow \infty$ in the finite horizon LQ control problem. The cost function becomes:

$$J(u, x_0) := \lim_{h\longrightarrow\infty} \left\{ \int_0^h (x^T C^T C x + u^T u)d\tau + x^T(h)Qx(h) \right\},\ Q \geq 0 \tag{43}$$

The minimizing control input can be shown to be given by $u = -B_2^T P x$ where P is the solution of the algebraic Riccati equation [35]:

$$PA + A^T P - PB_2 B_2^T P + C^T C = 0 \tag{44}$$

In the interest of brevity, the optimal attacks that maximize the LQ cost (43) will not be explicitly stated but follow simplified arguments similar to those presented in Section III. From (22) the optimal finite energy sensor attack $\Delta_y \in L^2([0,\infty), \mathbb{R}^n)$ with $\|\Delta_y\|_2 \leq M$ can be achieved by sensor attack signals of the form:

$$\Delta_y(t) = Mv_1\alpha(t),\ \ t > 0 \tag{45}$$

where v_1 is the normalized right singular vector corresponding to the maximal singular value of $B_2^T P$, and $\alpha(\cdot)$ is any positive scalar function in $L^2((0,\infty), \mathbb{R})$ of unit L^2-norm, i.e., $\|\alpha(t)\|_2 = 1$. At time $t = 0$ from (20) the sensor attack is given by letting $\Delta_y(0) = M\nu_1 - x(0)$, where ν_1 is the normalized eigenvector of P corresponding to the maximal eigenvalue of P, $\lambda_{\max}(P)$.

In the infinite horizon case for bounded sensor attack signals, $\Delta_y \in L^\infty([0,\infty), \mathbb{R}^n)$, with $\|\Delta\|_\infty \leq M$, from (38) it can be seen that the optimal attack signal is:

$$\Delta_y(t) = Mv_1,\ t > 0 \tag{46}$$

In the infinite horizon case, the singular vector v_1 is not function of time anymore. The attack signal (46) results in "infinitely" poor performance in the sense that the cost function (40), with $h = \infty$, satisfies:

$$\sup_{\|\Delta_y\|_\infty \leq M} J^\star(u, x_0, \infty, Q, \Delta_y) = \infty \tag{47}$$

V. Application to a Power Network

This section provides a numerical experiment to demonstrate the effectiveness of the proposed attack strategies and validate the theoretical analysis. A Power Model Network with a distributed control will be used for the simulation. We assume the attacker has access to the system parameters and the optimal sensor attack for the infinite horizon LQ problem (45) is simulated. First, the distributed controller for the power network is introduced.

A. Distributed Control for Power Networks

The control design for each generator in the power grid is necessary for the stability of the whole grid. Generally speaking, the optimal controller for the grid can be computed using a Linear Quadratic Regulator (LQR) controller. We propose to employ the distributed control scheme introduced by [36], for each generator. As noted in [36], it's valid to apply the control technique to this class of power units. For simplicity, we assume the connection topology to be a straight line (5 generators in a string). We consider the power grid as an interaction graph G in which each node represents a generation site and each edge denotes a transmission line. Assume each site has one synchronous generator and all generators are the identical. If sites i and k are adjacent, it is denoted by $i \sim k$.

Assume each generator supplies a time-varying current, time-varying power,and a constant voltage. The state of a single generator is represented by rotor rotation angle δ and described by the following swing equation [37]:

$$M\ddot{\delta} + D\dot{\delta} = P_m - P_e \tag{48}$$

where P_m is the mechanical power and P_e is the electrical power. M is the rotor inertia constant and D is the mechanical damping constant.

Assume all N generators operate near a stable point δ_0 and there is only small perturbation. Then, define $\epsilon_i = \delta_i - \delta_0$, based on the Kirchoff's current law and following the analysis in [38], (48) can be written as

$$\begin{aligned} \dot{\epsilon}_i &= \theta_i \\ \dot{\theta}_i &= -\frac{D}{M}\theta_i + \frac{1}{M}\sum_{k\sim i}\frac{V^2 X}{|Z|^2}(\epsilon_i - \epsilon_k) + \frac{1}{M}u_i \end{aligned} \tag{49}$$

where $Z = R + jX$ is the impedance of the transmission line. The parameter values are $M = 3.4$ for the rotor inertia,

$D = 0.8$ for the mechanical damping constant, $Z = 8+2*j$ as the impedance of the transmission line, and $V = 6.5$ as the generator voltage. Using the single generator equation (49), the state space equation for each generator is given as:

$$\dot{\mathbf{x}}_i = \mathbf{A}\mathbf{x}_i + \mathbf{B}u_i + \sum_{k=1}^{N} \mathbf{C}_k \mathbf{x}_k \tag{50}$$

where $\mathbf{x}_i = [\epsilon_i \quad \theta_i]^T$, and

$$\mathbf{A} := \begin{bmatrix} 0 & 1 \\ \frac{V^2 X}{M|Z|^2} & -\frac{D}{M} \end{bmatrix}, \ \mathbf{B} := \begin{bmatrix} 0 \\ \frac{1}{M} \end{bmatrix} \tag{51}$$

Let n_i be the number of generators that are adjacent to generator i, if $k \sim i$ then

$$\mathbf{C}_i := \begin{bmatrix} 0 & 0 \\ \frac{(n_i-1)V^2 X}{M|Z|^2} & 0 \end{bmatrix}, \ \mathbf{C}_k = \mathbf{0}_{2\times 2}, \ \text{otherwise} \tag{52}$$

This expression represents the effect on each local generator by states from other generators connected to it. The distributed controller employed will utilize this state information only. A suboptimal LQR controller for each generator is computed following [36].

B. Numerical Experiments

We consider the power network consisting of five generators outlined in the previous section as the example. The generators are connected in series and all the system parameters are fixed for this system.

1) Optimal Attacks for Infinite Horizon LQ: In this section the worst case attack for the LQ controller case with infinite horizon is simulated. The numerical results are shown in Fig. 1 - Fig. 3.

It should be noted that there are total 10 states for 5 generators, since each generator has two state variables. Fig. 1 shows the attack free system response. Based on the discussion in Section IV, (45) is the optimal attack signal from the attacker's viewpoint. The finite energy attack signal is with the bound $\|\Delta_y\|_2 \leq M$ where M is fixed to be 2 and the decay function was chosen as $\alpha(t) = 1/(t+1)$. Fig. 2 shows the states response during the optimal attack. As we can notice from Fig. 2, the states are significantly affected with the optimal attack. At the same time, the states asymptotically decay to 0 as the attack is a finite energy signal.

Fig. 3 shows the sum of all ten states for the attack free system and the system under optimal finite energy attack. These two comparison figures imply that the optimal sensor attack in (45) can affect significantly the performance of the system.

VI. CONCLUSIONS

In this work vulnerabilities under malicious attacks on the sensor signals attacks for observer-based and LQ controlled systems are analyzed. In particular, the error signals between states of attack free systems and systems subject to these attacks are quantified. Optimal sensor signal attacks for the finite and infinite horizon LQ control in terms of

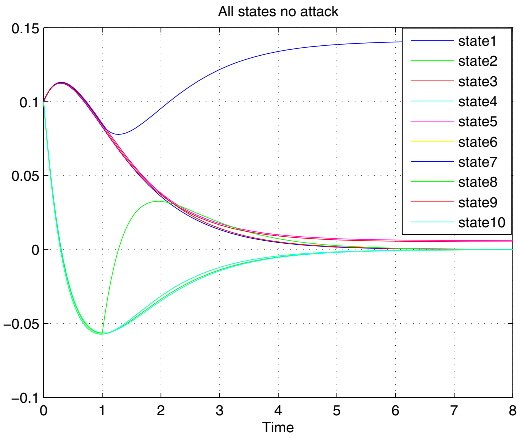

Fig. 1. All states without any attack.

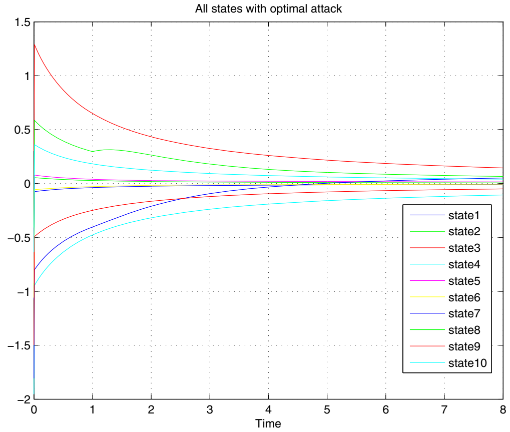

Fig. 2. All states with optimal attack.

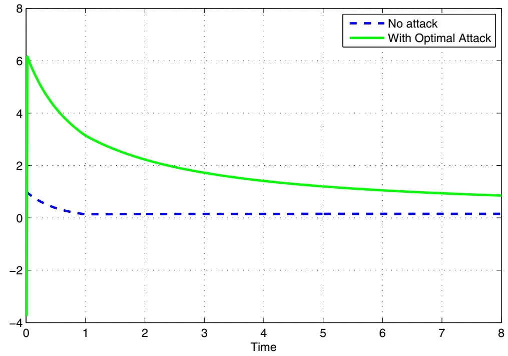

Fig. 3. Output comparison between attack free and optimal attack.

maximizing the corresponding cost functions are computed. Expressions for closed-loop systems under optimal signal attacks are provided. Illustrative numerical examples are provided showing the effect of cyber-attacks on controlled systems, in particular an application to a power network with distributed LQ controllers is provided. Future work includes determining optimal signal attacks for the linear quadratic Gaussian and H^∞ controllers where external disturbances and random noise come into play.

REFERENCES

[1] A. A. Cardenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," in *3rd USENIX workshop on Hot Topics in Security (HotSec '08), Associated with the 17th USENIX Security Symposium, San Jose, CA*, 2008.

[2] A. A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Workshop on Future Directions in Cyber-physical Systems Security, Newark, NJ*, 2009.

[3] J. Slay and M. Miller, "Lessons learned from the maroochy water breach," *Critical Infrastructure Protection*, 2007.

[4] P. Quinn-Judge, "Cracks in the system," *TIME Magazine*, 2002.

[5] J. Leyden, "Teen derails tram after hacking train network," *The Register*, 2008.

[6] A. Greenberg, "Hackers cut cities' power." *Forbes*, 2008.

[7] T. Greene, "Experts hack power grid in no time," *NetworkWorld*, 2008.

[8] D. Halperin, T. Heydt-Benjamin, B. Ransford, S. Clark, B. Defend, W. Morgan, T. K. K. Fu, and W. Maisel, "Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses," in *IEEE Symposium on Security and Privacy*, 2008.

[9] G. I. Security, "Tva needs to address weaknesses in control systems and networks," Tech. Rep. GAO-08-526, Report to Congressional Requesters, Tech. Rep., 2008.

[10] N. Falliere, L. O. Murchu, and E. Chien, "W32.stuxnet.dossier," http://www.symantec.com/content/en/us/enterprise/media/security response/whitepapers/w32 stuxnet dossier.pdf, 2011.

[11] M. F. P. Services and M. Labs., "Global energy attacks: Night," http://www.mcafee.com/us/resources/whitepapers/wp-global-energy-cyberattacks-night-dragon.pdf, 2011.

[12] "Kaspersky lab and itu research reveals new advanced cyber threat," http://www.kaspersky.com/about/news/virus/2012/Kaspersky Lab and ITU Research Reveals New Advanced Cyber Threat, 2012.

[13] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *The 28th International Conference on Distributed Computing Systems Workshops*, 2008.

[14] J. Eisenhauer, P. Donnelly, M. Ellis, and M. O'Brien, "Roadmap to secure control systems in the energy sector," *Energetics Incorporated. Sponsored by the U.S. Department of Energy and the U.S. Department of Homeland Security*, 2006.

[15] A. Greenberg, "America's hackable backbone," *Forbes*, 2007.

[16] U. Office, "Critical infrastructure protection. multiple efforts to secure control systems are under way, but challenges remain. technical report gao-07-1036," Technical Report GAO-07-1036, Tech. Rep., 2007.

[17] A. Melin, E. Ferragut, D. Fugate, R. Kisner, and J. Laska, "A mathematical framework for the analysis of cyber-resilient control systems," in *International Conf. on Resilient Control Systems*, 2013.

[18] A. M. Melin, R. Kisner, D. Fugate, and T. McIntyre, "Minimum state awareness for resilient control systems under cyber-attack," in *Future of Instrumentation International Workshop (FIIW)*, 2012.

[19] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems part i: Models and fundamental limitations," *arXiv:1202.6144v2 [math.OC]*, 2012.

[20] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber-physical system security for the electric power grid," *Proceedings of the IEEE, vol. 99, no. 1, pp. 1-15*, 2012.

[21] A. A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Workshop on Future Directions in Cyber-physical Systems Security, Newark, NJ*, 2009.

[22] H. Li, L. Lai, S. Djouadi, , and X. Ma, "Key establishment via common state information in networked," in *Proceedings of the American Control Conference, San Francisco, CA*, 2011.

[23] C. L. DeMarco, J. V. Sariashkar, and F. Alvarado, "The potential for malicious control in a competitive power systems environment," in *IEEE Int. Conf. on Control Applications, Dearborn, MI*, 1996.

[24] S. Amin, A. Cardenas, and S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," *Hybrid Systems: Computation and Control*, 2009.

[25] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *ACM Conference on Computer*, 2009.

[26] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *IEEE Conference on Decision and Control, Atlanta, GA*, 2010.

[27] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Conference on Communications, Control and Computing, Monticello, IL*, 2010.

[28] R. Smith, "A decoupled feedback structure for covertly appropriating network control systems," in *IFAC World Congress, Milan, Italy*, 2011.

[29] M. Zhu and S. Martinez, "Stackelberg-game analysis of correlated attacks in cyber-physical systems," in *American Control Conference,San Francisco, CA*, 2011.

[30] H. Li, L. Lai, and S. Djouadi, "Combating false reports for secure networked control in smart grid," in *Proc. of the IEEE International Conference in Communications (ICC), Kyoto, Japan*, 2011.

[31] C. G. Rieger, D. I. Gertman, and M. A. McQueen, "Resilient control systems: Next generation design research," in *Conference on Human Systems Interactions, Idaho Falls, ID*, 2009.

[32] Q. Zhu and T. Basar, "Robust and resilient control design for cyber-physical systems with an application to power systems," in *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC) Orlando, FL*, 2011.

[33] S. Zheng, T. Jiang, and J. Baras, "Robust state estimation under false data injection in distributed sensor networks," in *IEEE Globecom*, 2010.

[34] K. Zhou, J. Doyle, and K. Glover, *Robust and Optimal Control*. Prentice Hall, 1995.

[35] M. Green and D. Limebeer, *Linear Robust Control*. Prentice Hall, 1995.

[36] F. Borrelli and T. Keviczky, "Distributed lqr design for identical dynamically decoupled systems," *IEEE Trans. Automatic Control*, 2008.

[37] J. Machowski, J. W. Bialek, and J. R. Bumby, *Power System Dynamics: Stability and Control*. Wiley, 2008.

[38] H. Li and Z. Han, "Synchronization of power networks without and with communication infrastructur," in *IEEE SmartGridComm*, 2011.