Combating False Reports for Secure Networked Control in Smart Grid via Trustiness Evaluation

Husheng Li, Lifeng Lai and Seddik M. Djouadi

Abstract—Smart grid, equipped with modern communication infrastructures, is subject to possible cyber attacks. Particularly, false report attacks which replace the sensor reports with fraud ones may cause the instability of the whole power grid or even result in a large area blackout. In this paper, a trustiness system is introduced to the controller, who computes the trustiness of different sensors by comparing its prediction, obtained from Kalman filtering, on the system state with the reports from sensor. The trustiness mechanism is discussed and analyzed for the Linear Quadratic Regulation (LQR) controller. Numerical simulations show that the trustiness system can effectively combat the cyber attacks to smart grid.

I. INTRODUCTION

In recent years, smart grid has attracted significant interest in both communities of communications and power systems [3]. In a smart grid, modern communication technologies are used to convey information like system parameters (voltage, frequency, harmonics, etc) and power consumption information in order to improve the robustness, agility and efficiency of power grid. For example, as illustrated in Fig. 1, phasor measurement units (PMUs) send report to the power plant which takes actions to stabilize the power grid.

However, the communication infrastructure also brings vulnerability to the smart grid. An attacker can attack the communication links using various approaches such as jamming in the physical layer and Byzantine attack in the upper layers. The attack could result in delay or drop of report packets. The attacker may revise the report such that the received report is wrong, thus possibly incurring system instability and even large area blackout which brings the loss of millions of dollars. Therefore, a secure design of smart grid is in a pressing need.

In this paper, we study the trustiness framework based secure control in smart grid. We assume that each report could be substituted by a false report from an attacker [4]. Instead of studying the security protocols in the communication networks, which has been intensively studied for data networks like Internet, we focus on the controller side, i.e., the power plant, which is aware of possible attacks. The controller can predict the future system state and then evaluate the trustiness of reports from different PMUs. Based on these trustiness, the controller takes a corresponding control strategy, e.g., dropping untrusted packets. The trustiness will also be fed back to the PMUs such that the PMUs can adjust their security setups (e.g., the keys or the cryptography protocols). The system is illustrated in Fig. 1. For example, the reports from PMU A is attacked by the attacker. Then, the trustiness of the reports from PMU A is reduced, which may make the controller drop the corresponding reports. Meanwhile, PMU A may change its key or use a more secure protocol when it finds out that its trustiness has been significantly decreased.

Note that the reliability issues for control systems have been considered in supervisory control and data acquisition (SCADA) standard [10]. However, SCADA is mostly focused on the reliability subject to random failures, instead of malicious attacks. There are some studies on the control systems subject to malicious attacks [2] [4] [5]. [4] and [5] introduce general problems and approaches for secure control without exploring the details. [2] is focused on the control system subject to denial-of-service (DoS) attacks, thus mainly addressing the packet delays or losses. There have not been many studies on combating the proofing attacks, particularly, applying the trustiness framework in the secure control.

Note that the trustiness system has been widely used in information systems, e.g., [6] [11]. However, they are not designed for control systems. The unique system dynamics of the control system can be exploited to build the corresponding effective trustiness. In this paper, we apply Kalman filtering [2] to predict the system state using different combinations of PMU reports, thus realizing a cross-check of the report trustiness.

The remainder of this paper is organized as follows. The
system model of the controller and communication infrastructure is introduced in Section II. A mechanism of evaluating the trustiness of different PMUs is proposed in Section III. The numerical simulation results and conclusions are provided in Sections IV and V.

II. SYSTEM MODEL

In this section, we first introduce the linear system model for power grid. Then, we explain the LQR criterion of the control.

A. Linear System

We model the dynamics of power grid as a discrete-time linear system\(^2\) whose dynamics are given by

\[
\begin{align*}
    x(t+1) &= Ax(t) + Bu(t) + w(t), \\
    y(t) &= Cx(t) + n(t),
\end{align*}
\]

where \(x\) is an \(N\)-vector and represents the system state of the power grid, \(y\) is an \(M\)-vector representing the observations and \(u\) is the action taken by the controller. The matrices \(A, B\) and \(C\) are specified by the system. Both \(w(t)\) and \(n(t)\) are Gaussian noise. For simplicity, we suppose that each dimension of the observation vector \(y\) is sensed by a PMU. It is easy to extend to the general case in which the sensor reports have overlaps. Each PMU sends its observations to the controller via a communication channel since they are not located at the same place as the controller.

We use the following assumptions throughout the paper:

• Each report can be successfully received by the controller if there is no attack. This is reasonable for communication channels with good qualities. For the practical case of occasional packet drop, the control strategy can be obtained by considering the corresponding element in the observation matrix \(C\) as zero.

• Each report could be replaced with a false report by an intervening attacker. The attacker could intercept the report and insert its own one. However, we assume that not all reports are replaced.

• For simplicity, we assume that there is at most one attacker. The principle of the trustiness system can be extended to the case of multiple attackers at the cost of more computational cost.

B. LQR Control

When there is no attacker, we assume that the controller adopts the LQR control\(^7\) with an infinite time horizon with the cost function given by

\[
J = E \left[ \sum_{t=1}^{\infty} \beta^t \left( x^T(t)Qx(t) + u^T(t)Pu(t) \right) \right],
\]

where \(Q\) and \(P\) are positive definite matrices. The physical meaning of the objective function is given below:

\[\beta = \frac{1}{1 + \lambda},\]

\(\lambda\) is a strictly positive constant.

The term \(x^T(t)Qx(t)\) is the norm (with respect to the positive definite matrix \(Q\)) of the system state vector. The term \(u^T(t)Pu(t)\) is the norm of the action vector with respect to the matrix \(P\), which represents the cost due to the action itself.

Based on the cost function in (2), the LQR action \(u(t)\) is given by

\[
u(t) = -L\dot{x}(t),
\]

where \(\dot{x}(t)\) is the estimation of the system state fed back from a state estimator which will be explained later, and

\[
L = (B^TSA)^{-1}B^TSA,
\]

and the matrix \(x\) satisfies the algebraic Riccati Equation, which is given by

\[
S = A^T \left( S - SB \right) - \left( B^TSB + P \right)^{-1} B^TSA + Q.
\]

III. TRUSTINESS EVALUATION

In this section, we propose a mechanism for evaluating the trustiness of each sensor. The essential reason that the controller can evaluate the trustiness of each sensor is that the controller can predict the future state with some uncertainty. If the report from a sensor is significantly deviated from the prediction, then the controller can consider this sensor as unreliable and ignores its reports. The basic principle is to evaluate the trustiness of each PMU by comparing its report with the prediction obtained from the reports of other \(N-1\) PMUs. The procedure is illustrated in Fig. 2.

A. System State Prediction

In each time slot, the controller can predict the system state in the next time slot, i.e., \(\hat{x}(t+1)\), according to its own action \(u(t)\) and the current system state \(x(t)\). Since it is possible that there is one attacker, the controller computes \(N\) predictions by excluding one sensor in each prediction.

The Kalman filtering can be applied for the system state estimation. According to [9], the system state \(x(t+1)\) is

\[
\begin{align*}
    \hat{x}(t+1) &= \hat{x}(t) + \text{Kalman gain} \cdot (y(t) - C\hat{x}(t)), \\
    \text{Kalman gain} &= P(t)C^T \left( C \cdot P(t)C^T + Q \right)^{-1}, \\
    P(t+1) &= \text{Kalman gain} \cdot C \cdot P(t) + A \cdot P(t) \cdot A^T - \text{Kalman gain} \cdot C \cdot P(t) \cdot C^T \cdot \text{Kalman gain}. \\
\end{align*}
\]

\[\hat{x}(t+1)\] is the prediction obtained from the reports of other PMUs.
Gaussian distributed. When the observation \( y_n \) from PMU \( n \) is excluded, the expectation of the system state is given by

\[
x^n(t+1|t) = A_n x^n(t|t),
\]

where

\[
x^n(t|t) = x^n(t|t-1) + K^n(t) (y - C^n x(t|t-1)),
\]

where \( C^n \) is obtained by removing the \( n \)-th row from the observation matrix \( C \), and

\[
K^n(t) = \Sigma(t|t-1) \left( C^n \Sigma(t|t-1) (C^n)^T + \sigma_n^2 I \right)^{-1},
\]

and covariance matrix given by

\[
\Sigma^n(t|t) = \Sigma^n(t|t-1) - K^n C^n(t) \Sigma(t|t-1),
\]

where

\[
x^n(t+1|t) = A_n \Sigma^n(t|t) A^T + BQB^T.
\]

B. Trustiness Computation

Using the prediction obtained from the reports of PMUs except PMU \( n \), the \textit{a posteriori} probability of \( y_n(t) \), i.e., \( P(y_n(t)|y_{-n}(0: t)) \) (here \( y_{-n} \) means the observations excluding that of PMU \( n \)), is Gaussian distributed with the expectation given by

\[
u^n(t) = c_n x^n(t|t),
\]

where \( c_n \) is the \( n \)-th row in matrix \( C \), and variance

\[
\sigma^n(t) = c_n \Sigma^n(t|t) c_n^T.
\]

We denote by \( T_n \) the type of PMU \( n \). \( T_n = A \) if PMU \( n \) is an attacker; otherwise \( T_n = H \). Then, we define the suspicious level of PMU \( n \) as the following conditional probability, which is given by

\[
\pi_n(t) \triangleq P(T_n = A|y(0: t)).
\]

The trustiness can be defined as \( 1 - \pi_n(t) \).

The challenges for computing the suspicious level is the unknown attacking strategy.\footnote{It is possible that an attacker is captured and the strategy is known to the defender side. However, such an assumption is too strong for most systems.} We first assume that there must be an attacker. Using the Bayesian rule, we have

\[
\pi_n(t) = \frac{P(y(0: t)|T_n = A) \pi_n(t)}{\sum_{m=1}^{N} P(y(0: t)|T_m = A) \pi_m(t)},
\]

where the last approximation is obtained by decomposing the joint distribution into the product of the probabilities in each time slot. Note that this approximation is not rigorous. However, it simplifies the computation and the validity will be demonstrated in the numerical simulations. We do the following further simplification:

\[
P(y(s)|T_n = A) \approx \prod_{k}\left[ P(y_k(s)|T_k = H) \right],
\]

by assuming the independence among the PMUs. Although this assumption does not hold, it simplifies the analysis. We further assume that \( P(y_n(s)|T_n = A) \) is a constant since we have no knowledge about the attacker’s strategy. Substituting the approximation in (15) into (14), we obtain

\[
\pi_n(t) \approx \frac{1}{L + \sum_{m=1}^{N} \prod_{s} P(y_m(s)|T_m = H)}.
\]

We then approximate the probability \( P(y_n(s)|T_n = H) \) by \( N(u^n(t), \sigma^n(t)) \).

When it is possible that there is no attacker, it is easy to repeat the above procedure and obtain the following approximation for the suspicious level of PMU \( n \), which is given by

\[
\pi_n(t) \approx \frac{1}{L + \sum_{m=1}^{N} \prod_{s} P(y_m(s)|T_m = H)}.
\]

where \( L \) represents the \textit{a priori} likelihood that there is no attacker, which is given by

\[
L = \frac{P(\text{there is no attacker})}{P(\text{there is one attacker})},
\]

Obviously, the large \( L \) is, the less sensitive the controller is to possible attackers.

C. Secure Control Based on Trustiness

One approach to handle the attacker is to omit its reports once determining that it is an attacker. Alternatively, we propose a heuristic approach for the control based on the trustiness values, called \textit{weighted prediction}. Suppose that we still use the LQR control. Then, the control action taken at time \( t \) is given by

\[
u(t) = -L \hat{x}(t, \{P(T_n = A|y)\}_n),
\]

where the system state estimation is a function dependent on the suspicious levels of different PMUs. We set

\[
\hat{x}(t, \{P(T_n = A|y)\}_n) = \frac{\sum_{n=1}^{N} P(T_n = A|y) \hat{x}^n(t)}{\sum_{n=1}^{N} P(T_n = A|y)},
\]

where the estimation is the weighted sum of the system state estimations of different excluded PMUs. When the suspicious level of PMU \( n \) is high, the corresponding system state estimation \( \hat{x}^n(t) \), which excludes the reports from PMU \( n \), will dominate (recall that \( x^n(t) \) is obtained by excluding the reports from PMU \( n \)).

D. Algorithm Summary

The proposed algorithms are summarized in Procedure [1]
where we ignore the step \( \Delta t \). The details of the model can be found in [1].

We adopt the linear model analyzed in Example 6.2 of [1], in which the system is described using the following continuous-time linear dynamics:

\[
\dot{x}(t) = -M^{-1}Kx - M^{-1}u + w,
\]

(21)

where the matrix \( M \) is given by

\[
M = \begin{pmatrix}
2.1 & 1.55 & 1.55 & 0 & 0 & 0 & 0 \\
1.55 & 1.651 & 1.55 & 0 & 0 & 0 & 0 \\
1.55 & 1.55 & 1.605 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2.04 & 1.49 & 0 & 0 \\
0 & 0 & 0 & 1.49 & 1.526 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1786.9 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix},
\]

and the matrix \( K \) is given in (21) (at the top of the next page). The details of the model can be found in [1].

Since we discuss the discrete-time model in this paper, we approximate the continuous-time model by setting a small time step \( \Delta t \), which is given by

\[
x((n+1)\Delta t) - x(n\Delta t) \\
\approx -M^{-1}Kx(n\Delta t) - M^{-1}u(n\Delta t) + w(n\Delta t).
\]

(22)

Therefore, we assume the following discrete-time model:

\[
x(n+1) = (I - \Delta tM^{-1}K)x(n) - \Delta tM^{-1}u(n) + \Delta tw(n),
\]

(23)

where we ignore the step \( \Delta t \) in the index.

We assume that the PMUs can observe the system state directly, i.e., \( C = I \), each PMU for one dimension. We further assume that PMU 1 is malicious while all other PMUs are honest.

IV. NUMERICAL SIMULATIONS

In this section, we use numerical simulations to demonstrate the proposed trustiness system in smart grid.

A. Linear Model

We adopt the linear model analyzed in Example 6.2 of [1], in which the system is described using the following continuous-time linear dynamics:

\[
\dot{x}(t) = -M^{-1}Kx - M^{-1}u + w,
\]

(21)

where the matrix \( M \) is given by

\[
M = \begin{pmatrix}
2.1 & 1.55 & 1.55 & 0 & 0 & 0 & 0 \\
1.55 & 1.651 & 1.55 & 0 & 0 & 0 & 0 \\
1.55 & 1.55 & 1.605 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2.04 & 1.49 & 0 & 0 \\
0 & 0 & 0 & 1.49 & 1.526 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1786.9 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix},
\]

and the matrix \( K \) is given in (21) (at the top of the next page). The details of the model can be found in [1].

Since we discuss the discrete-time model in this paper, we approximate the continuous-time model by setting a small time step \( \Delta t \), which is given by

\[
x((n+1)\Delta t) - x(n\Delta t) \\
\approx -M^{-1}Kx(n\Delta t) - M^{-1}u(n\Delta t) + w(n\Delta t).
\]

(22)

Therefore, we assume the following discrete-time model:

\[
x(n+1) = (I - \Delta tM^{-1}K)x(n) - \Delta tM^{-1}u(n) + \Delta tw(n),
\]

(23)

where we ignore the step \( \Delta t \) in the index.

We assume that the PMUs can observe the system state directly, i.e., \( C = I \), each PMU for one dimension. We further assume that PMU 1 is malicious while all other PMUs are honest.

B. Evolution of Suspicious Level

In Figures 3 and 4 the evolution of the suspicious level is shown for the attacker and two honest PMUs. In Fig. 3, we assume that the report of the attacker is a Gaussian random variable with zero mean and variance 0.1. The attacker decides to attack or not to attack with probability 0.2 (called attack frequency). We observe that there is some fluctuation at the beginning. Then, at the times when the attacker launches attacks, the suspicious level of the attacker increases significantly. After around 80 time slots, the attacker can be well distinguished from the two honest PMUs. In Fig. 4 we assume that the attacker attaches a Gaussian noise with zero mean and variance 0.1 to the observation. The attack frequency is increased to 0.5. We observe that the suspicious level increases more smoothly.

C. Detection Delay and False Alarm

In Fig. 5 we plot the cumulative distribution function (CDF) curves of the time when the controller claims that an attacker is detected. We assume that the controller claims the attacker when the suspicious level is larger than 0.7. The cases of false alarms are excluded (note that a false alarm is defined as
the event that a honest PMU is claimed to be an attacker). We change the attack frequencies from 0.1 to 0.3. The linear system runs for 200 time slots. We observe that, when the attack frequency is 0.1, the attacker is not detected within the 200 time slots in around 40% cases. When the attack frequency is increased, the attacker will be detected faster.

In Fig. 6, we plot the receiver operation characteristic (ROC) curves, which show the average delay of detection (excluding the false alarms) and false alarm rate. The attack frequencies are 0.2 and 0.4, respectively. Again, we use the threshold of 0.7 for the suspicious level. Obviously, the detection delay increases when the attack frequency becomes larger with a fixed false alarm rate.

D. Comparison of Cost

In Figures 7 and 8 the cost averaged over 2000 time slots and 100 realizations is shown. We assume that \( Q = I \) and \( P = 0.01I \) in (2), i.e., we pay much more attention to the norm of the system state. We compare the costs using the weighted system state estimation in Section III and the system without any counter measure for the attacker (i.e., full trust to each PMU).

In our test, we found that a small amplitude attack (e.g., the attacker uses the same attack as in Fig. 5) causes very small impact on the system. This is because that the Kalman filter has certain inherent robustness since the attack can be partially mitigated by observations from other PMUs. Therefore, we assume that the report of PMU1 is the sum of the original report and a strong noise with a large variance, which is called attack amplitude. In Fig. 7 we assume that the attack amplitude is 100 and change the attack probabilities. We observe that, as the attack probabilities increases, the average cost of the full trust case is significantly increased. Meanwhile, the total cost is decreased as the attack frequency increases. A possible reason is that a higher attack frequency may cause a more rapid degradation of the trustiness of the attacker. The total cost is also shown for different attack amplitudes in Fig. 8. Again, the total cost increases as the attack amplitude is increased in the full trust case. In a contrast to Fig. 7, the average cost of the weighted system state case is not a monotonic function of the attack amplitude. The reason could be: when the attack amplitude is small, the attacker causes little damage to the system; when the amplitude is large, the controller can detect the attacker early and avoid the cost in later time slots.

V. CONCLUSIONS

In this paper, we have analyzed the possible spoof attack on the smart grid system, in which the attacker can intercept and true report and send its faked report to the controller, which can cause severe damage to the power grid. We have proposed a trustiness system for the controller, in which multiple Kalman filtering processes, with each PMU excluded, are used to cross check the suspiciousness of each PMU. The suspicious levels are then used as weights for the system state estimation for the LQR control. Numerical simulations have shown that the attacker can be effectively detected and the weighted system prediction approach significantly outperforms the system unaware of possible attacks.

Our future work includes the following aspects:

- The multiple attacker case.
• The optimal attack strategy of the attacker against the proposed trustiness system.

REFERENCES

[1] P. M. Anderson and A. A. Fouad, *Power System Control and Stability*, 2nd edition, IEEE Press and Wiley-Interscience, 2003.

[2] S. Amin, A. A. Cárdenas and S. Sastry, “Safe and secure networked control systems under denial-of-service attacks,” *Lecture Notes in Computer Science*, Springer, 2009.

[3] ISO New England Inc., *Overview of the Smart Grid: Policies, Initiatives and Needs*, Feb. 17, 2009.

[4] A. A. Cárdenas, S. Amin and S. Sastry, “Secure control: Towards survivable cyber-physical systems,” in *Proc. of the 28th International Conference on Distributed Computing Systems Workshops*, 2008.

[5] A. A. Cárdenas, S. Amin and S. Sastry, “Research challenges for the security of control systems,” in *Proc. of the 3rd Conference on Hot Topics in Security*, 2008.

[6] A. Jswang, R. Ismail and C. Boyd, “A survey of trust and reputation systems for online service provision,” *Decision Support Systems*, vol. 43, pp.618–644, March 2007.

[7] H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems*, 1st Edition, Wiley-Interscience, 1972.

[8] A. S. Matveev and A. V. Savkin, *Estimation and Control Over Communication Networks*, Birkhuser, 2009.

[9] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd edition, Springer, 1994.

[10] R. Lemos, “SCADA system makers pushed toward security,” SecurityFocus, 2007.

[11] F. Stumpt, M. Benz, M. Hermanowski and C. Eckert, “An approach to a trustworthy system architecture using virtualization,” *Lecture Notes in Computer Science*, 2007.

[12] C. W. Taylor, D. C. Erickson, K. E. Martin, R. E. Wilson and V. Venkatasubramanian, “WACS-Wide-are stability and voltage control system: R&D and online demonstration,” *Proceedings of the IEEE*, pp.892–906, May 2005.

[13] D. Wei, Y. Lu, M. Jafari, P. Skare and K. Rohde, “An integrated security system of protecting smart grid against cyber attacks,” in *Proc. of Innovative Smart Grid Technologies Conference*, 2010.