Supplementary File

Jue Tian, Rui Tan, *Senior Member, IEEE,* Xiaohong Guan, *Fellow, IEEE,* Zhanbo Xu, *Member, IEEE,* and Ting Liu, *Member, IEEE*

This document includes the supplemental materials for the paper titled "Moving Target Defense to Detecting Stuxnet-like Attacks."

APPENDIX A
INSTANTIATED MODELS OF DYNAMIC STATE ESTIMATOR AND ANOMALY DETECTOR

This section introduces the Kalman Filter-based state estimator, the χ^2 anomaly detector, and the linear controller, which have been widely used in literature and real systems [1], [2]. The simulations conducted in Section V-B use these instantiated models. First, the Kalman Filter (KF) instantiates Eq. (3) as $\hat{\mathbf{x}}[k+1] = \mathbf{A}\hat{\mathbf{x}}[k] + \mathbf{B}\mathbf{u}[k] + \mathbf{F}(\mathbf{y}[k+1] - \mathbf{C}\mathbf{A}\hat{\mathbf{x}}[k] - \mathbf{C}\mathbf{B}\mathbf{u}[k])$, where $\mathbf{F} \in \mathbb{R}^{n\times m}$ is the Kalman gain matrix given by $\mathbf{F} = \mathbf{P}\mathbf{C}^T(\mathbf{C}\mathbf{P}\mathbf{C}^T + \mathbf{R})^{-1}$ and $\mathbf{P}$ is the solution of the equation $\mathbf{P} = \mathbf{A}\mathbf{P}\mathbf{A}^T + \mathbf{Q} - \mathbf{A}\mathbf{P}\mathbf{C}^T(\mathbf{C}\mathbf{P}\mathbf{C}^T + \mathbf{R})^{-1}\mathbf{C}\mathbf{P}\mathbf{A}^T$. In addition, Eq. (4) becomes $\hat{\mathbf{y}}[k+1] = \mathbf{C}\mathbf{A}\hat{\mathbf{x}}[k] + \mathbf{C}\mathbf{B}\mathbf{u}[k]$. Second, the χ^2 detector computes a scalar residual as $r[k] = \boldsymbol{\epsilon}[k]^T\mathbf{G}^{-1}\boldsymbol{\epsilon}[k]$, where $\mathbf{G} \in \mathbb{R}^{m\times m}$ is the covariance of $\boldsymbol{\epsilon}[k]$. When the KF-based state estimator is used, $\mathbf{G} = \mathbf{C}\mathbf{P}\mathbf{C}^T + \mathbf{R}$ and the $r[k]$ follows a χ^2_m distribution. The χ^2 detector yields a positive detection result if $r[k+1] > \chi^2_{m,\alpha}$, where $\chi^2_{m,\alpha}$ represents the $100\alpha\%$-percentile of the χ^2_m distribution. Thus, the χ^2 detector ensures a false alarm rate of $(1-\alpha)$. Third, in the simulations, we use a negative feedback linear controller of $\mathbf{u}[k] = -\mathbf{H}(\hat{\mathbf{x}}[k] - \mathbf{x}_0)$, where $\mathbf{H} \in \mathbb{R}^{l\times n}$ is a constant matrix.

APPENDIX B
DISCUSSION OF ATTACK CAPABILITIES

This section provides justification for the attacker's capabilities presented in Section III-A. Critical CPSes are often the target of strong attackers (such as insiders or hostile national rivals), who can obtain detailed knowledge of the system and the anomaly detection methods through various means. Besides, all traditional control systems adopt a centralized control theme. Specifically, all the sensor measurements are transmitted to a control center for system monitoring, and all the control commands are transmitted from the control center for system regulation. If the attackers can eavesdrop on and tamper with the data flows through several critical routers close to the control center or directly do so in the control center (which is not impossible as evidenced in the recent high-profile intrusions such as Stuxnet), they can compromise the control commands and sensor measurements. When the attackers can corrupt only a subset of the control commands or sensor measurements, they can still perform local stealthy SL attacks, according to the system topology (i.e., $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ mentioned in Section II-A). As the focus of this paper is to develop a defense approach, it is beneficial to assume a strong adversary model (i.e., assume the attackers can compromise all commands and measurements). The defense based on this strong adversary model will be also effective to the weaker attackers who can corrupt a subset of the commands or measurements only.

APPENDIX C
PROOF OF LEMMA 1

Proof. (Sufficiency) We prove it through the mathematical induction method. Since $\mathbf{y}_m[k] = \mathbf{y}[k]$, $\forall k \in \mathbb{N}$, we have $\mathbf{Y}_m[k] = \mathbf{Y}[k]$, $\forall k \in \mathbb{N}$. In the initial time period, since $\mathbf{Y}_m[0] = \mathbf{Y}[0]$, according to Eqs. (3) and (9), we have $\hat{\mathbf{x}}_m[0] = \hat{\mathbf{x}}[0]$, i.e., $\hat{\mathbf{X}}_m[0] = \hat{\mathbf{X}}[0]$; since $\mathbf{Y}_m[0] = \mathbf{Y}[0]$ and $\hat{\mathbf{X}}_m[0] = \hat{\mathbf{X}}[0]$, according to Eqs. (6) and (12), we have $\mathbf{u}_m[0] = \mathbf{u}[0]$, i.e., $\mathbf{U}_m[0] = \mathbf{U}[0]$. Supposing that in the kth time period, $\hat{\mathbf{X}}_m[k] = \hat{\mathbf{X}}[k]$ and $\mathbf{U}_m[k] = \mathbf{U}[k]$. In the $(k+1)$th time period, since $\mathbf{Y}_m[k+1] = \mathbf{Y}[k+1]$, $\hat{\mathbf{X}}_m[k] = \hat{\mathbf{X}}[k]$ and $\mathbf{U}_m[k] = \mathbf{U}[k]$, according to Eqs. (3) and (9), we have $\hat{\mathbf{x}}_m[k+1] = \hat{\mathbf{x}}[k+1]$. Then, $\hat{\mathbf{X}}_m[k+1] = \hat{\mathbf{X}}[k+1]$. Moreover, since $\mathbf{Y}_m[k+1] = \mathbf{Y}[k+1]$, $\hat{\mathbf{X}}_m[k+1] = \hat{\mathbf{X}}[k+1]$ and $\mathbf{U}_m[k] = \mathbf{U}[k]$, according to Eqs. (6) and (12), we have $\mathbf{u}_m[k+1] = \mathbf{u}[k+1]$. Then, $\mathbf{U}_m[k+1] = \mathbf{U}[k+1]$. Since both the basis and inductive step have been performed, we have $\hat{\mathbf{X}}_m[k] = \hat{\mathbf{X}}[k]$ and $\mathbf{U}_m[k] = \mathbf{U}[k]$, for $\forall k \in \mathbb{N}$. Therefore, according to Eqs. (4) and (10), we have $\hat{\mathbf{y}}_m[k] = \hat{\mathbf{y}}[k], \forall k \in \mathbb{N}$. Then, from $\mathbf{y}_m[k] = \mathbf{y}[k]$ and $\hat{\mathbf{y}}_m[k] = \hat{\mathbf{y}}[k]$, we have $\boldsymbol{\epsilon}_m[k] = \boldsymbol{\epsilon}[k]$, for $\forall k \in \mathbb{N}$, i.e., the attack is stealthy.

(Necessity) We prove it through the mathematical induction method. In the initial time period, since the attack is stealthy and $\hat{\mathbf{y}}_m[0] = \hat{\mathbf{y}}[0]$, we have $\mathbf{0} = \boldsymbol{\epsilon}_m[0] - \boldsymbol{\epsilon}[0] = \mathbf{y}_m[0] - \mathbf{y}[0]$. Then, according to Eqs. (3) and (9), we have $\hat{\mathbf{x}}_m[0] = \hat{\mathbf{x}}[0]$. Moreover, according to Eqs. (6) and (12), we have $\mathbf{u}_m[0] = \mathbf{u}[0]$. That is, $\mathbf{Y}_m[0] = \mathbf{Y}[0]$, $\hat{\mathbf{X}}_m[0] = \hat{\mathbf{X}}[0]$ and $\mathbf{U}_m[0] = \mathbf{U}[0]$. Supposing that in the kth time period, $\mathbf{Y}_m[k] = \mathbf{Y}[k]$, $\hat{\mathbf{X}}_m[k] = \hat{\mathbf{X}}[k]$ and $\mathbf{U}_m[k] = \mathbf{U}[k]$. In the $(k+1)$th time period, according to Eqs. (4) and (10), we have $\hat{\mathbf{y}}_m[k+1] = \hat{\mathbf{y}}[k+1]$. Since the attack is stealthy, we have $\mathbf{0} = \boldsymbol{\epsilon}_m[k+1] - \boldsymbol{\epsilon}[k+1] = \mathbf{y}_m[k+1] - \mathbf{y}[k+1]$, i.e., $\mathbf{Y}_m[k+1] = \mathbf{Y}[k+1]$. From $\mathbf{Y}_m[k+1] = \mathbf{Y}[k+1]$, $\hat{\mathbf{X}}_m[k] = \hat{\mathbf{X}}[k]$ and $\mathbf{U}_m[k] = \mathbf{U}[k]$, we have $\hat{\mathbf{x}}_m[k+1] = \hat{\mathbf{x}}[k+1]$ according

to Eqs. (3) and (9), i.e., $\hat{\mathbf{X}}_m[k+1] = \hat{\mathbf{X}}[k+1]$. Therefore, according to Eqs. (6) and (12), we have $\mathbf{u}_m[k+1] = \mathbf{u}[k+1]$, i.e., $\mathbf{U}_m[k+1] = \mathbf{U}[k+1]$. Since both the basis and inductive step have been performed, we have $\mathbf{Y}_m[k] = \mathbf{Y}[k]$, $\hat{\mathbf{X}}_m[k] = \hat{\mathbf{X}}[k]$ and $\mathbf{U}_m[k] = \mathbf{U}[k]$, for $\forall k \in \mathbb{N}$, according to the mathematical induction. That is, $\mathbf{y}_m[k] = \mathbf{y}[k]$, for $\forall k \in \mathbb{N}$. □

APPENDIX D
PROOF OF SCALING ATTACKS' STEALTHINESS

Proof. We prove it through the mathematical induction method. Initially, $\mathbf{b}[0] = \mathbf{0}$ and $\mathbf{Y}_m[0] = \mathbf{Y}[0]$. Assume $\mathbf{Y}_m[k] = \mathbf{Y}[k]$. Then, $\mathbf{U}_m[k] = \mathbf{U}[k]$. In the $(k+1)$th time period, $\mathbf{C}^+(\mathbf{y}_a[k] + \mathbf{b}[k]) = \mathbf{C}^+\mathbf{y}_m[k] = \mathbf{C}^+\mathbf{y}[k] = \mathbf{x}[k]$. Thus, $\mathbf{C}^+\mathbf{b}[k] = \mathbf{x}[k] - \mathbf{x}_a[k]$. Moreover, $\mathbf{y}_a[k+1] = \mathbf{C}\mathbf{A}\mathbf{x}_a[k] + (1+\lambda_k)\mathbf{C}\mathbf{B}\mathbf{u}[k]$. Then, $\mathbf{y}_a[k+1] - \mathbf{C}\mathbf{A}\mathbf{x}_a[k] = (1+\lambda_k)\mathbf{C}\mathbf{B}\mathbf{u}[k]$. Hence, $\mathbf{b}[k+1] = -\lambda_k\mathbf{C}\mathbf{B}\mathbf{u}[k] + \mathbf{C}\mathbf{A}(\mathbf{x}[k] - \mathbf{x}_a[k])$. Then, $\mathbf{y}_m[k+1] = \mathbf{C}\mathbf{A}\mathbf{x}_a[k] + (1+\lambda_k)\mathbf{C}\mathbf{B}\mathbf{u}[k] - \lambda_k\mathbf{C}\mathbf{B}\mathbf{u}[k] + \mathbf{C}\mathbf{A}(\mathbf{x}[k] - \mathbf{x}_a[k]) = \mathbf{y}[k+1]$. Therefore, $\mathbf{Y}_m[k+1] = \mathbf{Y}[k+1]$. From Lemma 1, the attack is stealthy. □

APPENDIX E
FREQUENTLY STUDIED ATTACKS IN LITERATURE EXCEPT FOR SL ATTACKS

This section briefly introduces several frequently studied attacks in the literature. We note that these attacks are not stealthy in the dynamic state estimator and anomaly detector, and thus not the focus of this paper.

- Denial-of-service (DoS) attack: The DoS attack is a cyber-attack in which the attackers jam the communication channels and prevent the control signal or sensor measurement from reaching the destination. In a DoS attack, the attackers usually overwhelm the targeted devices or exhaust relevant resources by sending excessive messages [3]. Zhang et al. [4] studied the optimal attack strategies of DoS attack to degrade the system performance. Agah and Das [5] proposed a protocol based on game theory to prevent the DoS attack.
- False data injection attack: FDI attacks can be launched against the state estimation of power grids while keeping stealthy to the SE's BDD mechanism [6], if the attackers know the details of the BDD and can compromise the sensor measurements. FDI attacks for different malicious objectives have been studied in [7], [8] and [9]. The stealthy FDI attacks can be prevented by protecting selected sensor measurements [10], or securing PMU data [11]. The attacks can be detected by using more adaptive algorithms [12], [13].
- Other attacks: Huang et al. [14] and Sridhar et al. [15] proposed the min-and-max attack, the scaling attack, the pulse attack and the random attack against either the sensors or actuators. However, according to the analysis in Section III-B, these attacks that compromise either the control signals or the sensor measurements are not stealthy.

APPENDIX F
SL ATTACK IMPLEMENTATION IN AN OFF-LINE MODE AND AN ILLUSTRATING CASE

This section separately analyzes whether the three SL attacks can be launched in an off-line mode.

- MISA: According to Definition 2, the attackers can design the command injection $\mathbf{a}$ in a priori, and calculate the measurement injection $\mathbf{b}$ for each time period. Since the attack vectors $\mathbf{a}$ and $\mathbf{b}$ are independent to the real-time system states or measurements, the MISA can be launched off-line.
- Scaling Attack: According to Definition 3, the measurement injection $\mathbf{b}$ is usually designed according to the whole real-time sensor measurements $\mathbf{y}_a$. Thus, the scaling attack cannot be launched off-line.
- Replay Attack: To launch the replay attack, the attackers should firstly judge whether the system has converged, by monitoring the control signals and the sensor measurements. In an off-line mode, this may need the communication between the injected malware on each router, but needs weak synchronization requirement. After convergence, according to Definition 4, the compromised measurement of each sensor will be replaced by its historical data, which can be achieved in a distributed manner by the malware on each router. Thus, the replay attack can be launched off-line.

1) A Numerical Case of MISA: To better illustrate the concept of the CPS loop and the SL attack, here we provide a numerical case of MISA's implementation. For simplicity, we consider a noiseless LTI system, which is given by (7) and (8) with the following system parameters:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{w} = \mathbf{0}, \mathbf{v} = \mathbf{0}.$$

The system state estimation process in Eqs. (9) and (10), and the control algorithm in Eq. (12) are given by:

$$\begin{aligned} \hat{\mathbf{x}}_m[k+1] &= \mathbf{y}_m[k+1], \\ \hat{\mathbf{y}}_m[k+1] &= \mathbf{A}\hat{\mathbf{x}}_m[k+1] + \mathbf{B}\mathbf{u}_m[k], \\ \mathbf{u}_m[k+1] &= 0.5(\mathbf{x}_0 - \hat{\mathbf{x}}_m[k+1]). \end{aligned}$$

In addition, the initial system state is $\mathbf{x}[0] = [10\ 10]^T$ and the desired target state is $\mathbf{x}_0 = [0\ 0]^T$. Due to noiseless, the threshold $\boldsymbol{\epsilon}_0$ can be set as $\mathbf{0}$, which ensures a zero false alarm rate.

TABLE I
NUMERICAL CASE OF MISA.

Time	$\mathbf{x}(\mathbf{y})$	$\mathbf{x}_a(\mathbf{y}_a)$	$\mathbf{b}$	$\mathbf{y}_m$	$\hat{\mathbf{y}}_m$	$\boldsymbol{\epsilon}_m$	$\mathbf{u}_m$	$\mathbf{a}$	$\mathbf{u}_a$
0	$[10\ 10]^T$	$[10\ 10]^T$	$[0\ 0]^T$	$[10\ 10]^T$	$[10\ 10]^T$	$[0\ 0]^T$	$[-5\ -5]^T$	$[0\ 0]^T$	$[-5\ -5]^T$
1	$[15\ 5]^T$	$[15\ 5]^T$	$[0\ 0]^T$	$[15\ 5]^T$	$[15\ 5]^T$	$[0\ 0]^T$	$[-7.5\ -2.5]^T$	$[0\ 0]^T$	$[-7.5\ -2.5]^T$
2	$[12.5\ 2.5]^T$	$[12.5\ 2.5]^T$	$[0\ 0]^T$	$[12.5\ 2.5]^T$	$[12.5\ 2.5]^T$	$[0\ 0]^T$	$[-6.25\ -1.25]^T$	$[5\ 5]^T$	$[-1.25\ 3.75]^T$
3	$[8.75\ 1.25]^T$	$[13.75\ 6.25]^T$	$[-5\ -5]^T$	$[8.75\ 1.25]^T$	$[8.75\ 1.25]^T$	$[0\ 0]^T$	$[-4.38\ -0.63]^T$	$[5\ 5]^T$	$[0.63\ 4.38]^T$
4	$[5.63\ 0.63]^T$	$[20.63\ 10.63]^T$	$[-15\ -10]^T$	$[5.63\ 0.63]^T$	$[5.63\ 0.63]^T$	$[0\ 0]^T$	$[-2.81\ -0.31]^T$	$[5\ 5]^T$	$[2.19\ 4.69]^T$
5	$[3.44\ 0.31]^T$	$[33.44\ 15.31]^T$	$[-30\ -15]^T$	$[3.44\ 0.31]^T$	$[3.44\ 0.31]^T$	$[0\ 0]^T$	$[-1.72\ -0.16]^T$	$[5\ 5]^T$	$[3.28\ 4.84]^T$

Table I shows the numerical result of MISA. The column "$\mathbf{x}(\mathbf{y})$" shows the system state in the absence of the attack, which rapidly converges to the desired target state $\mathbf{x}_0$. Unfortunately, the attacker designs the attack plan in advance, which is shown in the column "$\mathbf{a}$" and "$\mathbf{b}$". The attacker begins to inject the malicious control commands from the second time period. The column "$\mathbf{x}_a(\mathbf{y}_a)$" shows the real system state, which deviates from the desired target state $\mathbf{x}_0$ and has reached to $[33.44\ 15.31]^T$ in the fifth time period. Meanwhile, the attacker corrupts the sensor readings from the third time period to cover the ongoing attack. The residual error (shown in the column "$\boldsymbol{\epsilon}_m$") is always $[0\ 0]^T$ and thus the MISA is stealthy to the system. Note that the sensor reading in the presence of the attack is same as that in the absence of the attack, i.e., $\mathbf{y}_m = \mathbf{y}$, which is consistent with Lemma 1.

REFERENCES

[1] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *IEEE Allerton Conf. Communication, Control, and Computing*, 2009.

[2] S. Lakshminarayana, T. Z. Teng, D. K. Y. Yau, and R. Tan, "Optimal attack against cyber-physical control systems with reactive attack mitigation," in *ACM Intl. Conf. Future Energy Systems (e-Energy)*, 2017.

[3] M. McDowell, "Understanding denial-of-service attacks," *National Cyber Alert System, Cyber Security Tip ST04-015.2004*, 2004.

[4] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal denial-of-service attack scheduling with energy constraint," *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 3023–3028, 2015.

[5] A. Agah and S. K. Das, "Preventing dos attacks in wireless sensor networks: A repeated game theory approach." *IJ Network Security*, vol. 5, no. 2, pp. 145–153, 2007.

[6] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Info. & Syst. Security*, vol. 14, no. 1, 2011.

[7] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," in *First IEEE International Conference on Smart Grid Communications*, 2010, pp. 226–231.

[8] Y. Yuan, Z. Li, and K. Ren, "Modeling load redistribution attacks in power systems," *IEEE Trans. Smart Grid*, vol. 2, no. 2, 2011.

[9] J. Kim, L. Tong, and R. J. Thomas, "Dynamic attacks on power systems economic dispatch," in *Asilomar Conference on Signals, Systems and Computers*, 2015, pp. 345–349.

[10] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting false data injection attacks on dc state estimation," in *Preprints of the First Workshop on Secure Control Systems, CPSWEEK*, Stockholm, Sweden, 2010.

[11] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks: characterizations and countermeasures," in *SmartGridComm*, 2011.

[12] Y. Huang, H. Li, K. A. Campbell, and Z. Han, "Defending false data injection attack on smart grid network using adaptive cusum test," in *Annual Conf. Inf. Sci. & Syst.*, 2011.

[13] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Trans. Smart Grid*, vol. 5, no. 2, 2014.

[14] Y.-L. Huang, A. A. Cárdenas, S. Amin, Z.-S. Lin, H.-Y. Tsai, and S. Sastry, "Understanding the physical and economic consequences of attacks on control systems," *International Journal of Critical Infrastructure Protection*, vol. 2, no. 3, pp. 73–83, 2009.

[15] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 580–591, 2014.