Two-Way Coding and Attack Decoupling in Control Systems Under Injection Attacks

Song Fang
School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology
Stockholm, Sweden
sonf@kth.se

Karl Henrik Johansson
School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology
Stockholm, Sweden
kallej@kth.se

Mikael Skoglund
School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology
Stockholm, Sweden
skoglund@kth.se

Henrik Sandberg
School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology
Stockholm, Sweden
hsan@kth.se

Hideaki Ishii
Department of Computer Science, Tokyo Institute of Technology
Yokohama, Japan
ishii@c.titech.ac.jp

ABSTRACT

In this paper, we introduce the method of two-way coding, a concept originating in communication theory characterizing coding schemes for two-way channels, into feedback control systems under injection attacks. We propose the notion of attack decoupling, and show how the controller and the two-way coding can be co-designed to nullify the transfer function from attack to plant, rendering the attack effect zero both in transient phase and in steady state.

KEYWORDS

Cyber-physical system, networked control system, cyber-physical security, two-way channel, two-way coding

1 INTRODUCTION

The concept of two-way communication channels dates back to Shannon [13]. As its name indicates, in two-way channels, signals are transmitted simultaneously in both directions between the two terminals of communication. Accordingly, coding for two-way channels should make use of the information contained in the data transmitted in both directions; in other words, the coding schemes are also two-way, and thus are referred to as two-way coding [1]. Inherently, the communication channels in networked feedback control systems are two-way channels, with the controller side and the plant side being viewed as the two terminals of communication, respectively. Nevertheless, approaches based on two-way coding for the two-way channels in networked feedback systems are rarely seen in the literature. One exception is the so-called scattering transformation utilized in the tele-operation of robotics [6]; in a broad sense, scattering transformation can be viewed as a special class of two-way coding to resolve the issue of two-way time delays, the most essential characterization and the main issue of the two-way channels modeled on the input-output level in the problem of tele-operation.

Particularly in the cyber-physical security problems (see, e.g., [7, 9–12, 14, 15, 19] and the references therein) of networked control systems, to the best of our knowledge, only one-way coding has been employed. The authors of [18] introduced (one-way) encryption matrices into control systems to achieve confidentiality and integrity. In [8], the authors considered a method of coding (using one-way coding matrices) the sensor outputs in order to detect stealthy false data injection attacks in cyber-physical systems. Modulation matrices, which are one-way, were inserted into cyber-physical systems in [5] to detect covert attacks and zero-dynamics attacks. Dynamic one-way coding was applied to detect and isolate routing attacks [4] and replay attacks [3]. For remote state estimation in the presence of eavesdroppers, the so-called state-secrecy codes were introduced [16], which are also inherently one-way coding schemes. On the other hand, one-way coding has its inherent limitations; for instance, one-way coding in general cannot eliminate the unstable poles nor nonminimum-phase zeros of the plant nor the controller [2], which are most critical issues in the defense against, e.g., zero-dynamics attacks [15].

In our previous work [2], we examined how the presence of two-way coding in linear time-invariant (LTI) feedback control systems can make the zeros and/or poles of the equivalent plant as viewed by the attacker

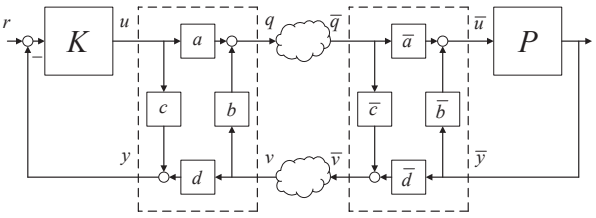

Figure 1: A networked feedback system with two-way coding.

all different from those of the original plant, and under some additional assumptions (i.e., the plant is stabilizable by static output feedback), the equivalent plant may even be made stable and/or minimum-phase. In the particular case of zero-dynamics attacks, it is then implicated that the attacks will be detected if designed according to the original plant, while the attack effect may be corrected in steady state if the attacks are to be designed with respect to the equivalent plant.

To prevent possible damages during the transient phase even when the attack affect can be corrected in steady state, in this paper we propose the notion of attack decoupling. For LTI systems, we say that a certain attack is decoupled if the transfer function from attack to plant input/output is made zero, without making zero the transfer function from reference to plant input/output. As such, when attack decoupling is achieved, the attack response will be zero both in transient phase and in steady state. We then examine conventional feedback systems, feedback systems with one-way coding, as well as feedback systems with two-way coding, and find that it is only in feedback systems with two-way coding that attacks in the uplink or downlink channels can be decoupled.

2 TWO-WAY CODING

Consider the single-input single-output (SISO) system depicted in Fig. 1. Herein, K denotes the controller while P denotes the plant. The reference signal is $r(t) \in \mathbb{R}$ and the plant output is $\overline{y}(t) \in \mathbb{R}$. In addition, let $u(t)$, $\overline{u}(t)$, $y(t)$, $q(t)$, $\overline{q}(t)$, $v(t)$, $\overline{v}(t) \in \mathbb{R}$.

Definition 2.1. The (static) two-way coding is defined as

$$\left[\begin{array}{c} q(t) \\ y(t) \end{array}\right] = M \left[\begin{array}{c} u(t) \\ v(t) \end{array}\right],\ M = \left[\begin{array}{cc} a & b \\ c & d \end{array}\right]. \quad (1)$$

Herein, $a, b, c, d \in \mathbb{R}$ are chosen such that

$$ad \neq 0,\ ad - bc \neq 0. \quad (2)$$

Strictly speaking, it should be further assumed that $|ad - bc| < \infty$.

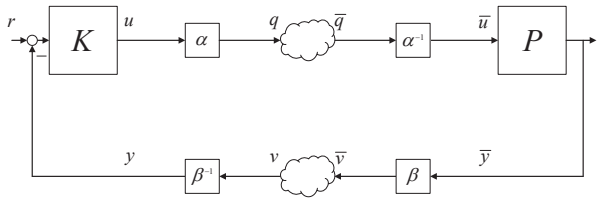

Figure 2: A networked feedback system with one-way coding.

Herein, two-way coding (operating in a feedback loop) represents a two-way transformation that takes in the signal in the forward path and the signal in the feedback path, and outputs a new signal to the forward path and a second new signal that passes on in the feedback path. In comparison, Fig. 2 depicts a system with one-way coding schemes, which are one-way transformations that either take in the signal in the forward path and output a new signal that passes on in the forward path, or input the signal in the feedback path and output a signal that continues in the feedback path; herein, $\alpha, \beta \in \mathbb{R}$ and $0 < |\alpha|, |\beta| < \infty$.

For simplicity, we denote the inverse of two-way coding M as

$$\left[\begin{array}{cc} \overline{a} & \overline{b} \\ \overline{c} & \overline{d} \end{array}\right] = M^{-1} = \left[\begin{array}{cc} \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right], \quad (3)$$

where $\overline{a}, \overline{b}, \overline{c}, \overline{d} \in \mathbb{R}$. As illustrated on the plant side in Fig. 1, the inverse of two-way coding M denotes another two-way coding.

3 ATTACK DECOUPLING

In this section, we analyze in particular LTI feedback control systems. Consider the SISO feedback system with two-way coding depicted in Fig. 3. Assume that herein the controller K and plant P are LTI with transfer functions $K(s)$ and $P(s)$, respectively. In addition, let $r(t)$, $u(t)$, $\overline{u}(t)$, $y(t)$, $\overline{y}(t)$, $q(t)$, $\overline{q}(t)$, $v(t)$, $\overline{v}(t) \in \mathbb{R}$. Meanwhile, suppose that injection (additive) attacks $w(t) \in \mathbb{R}$ and $z(t) \in \mathbb{R}$ exist in the forward path and feedback path of the control systems, respectively. Let $R(s), U(s), \overline{U}(s), Y(s), \overline{Y}(s), Q(s), \overline{Q}(s), V(s), \overline{V}(s)$, $W(s)$, $Z(s)$ represent the Laplace transforms, assuming that they exist, of the signals $r(t)$, $u(t)$, $\overline{u}(t)$, $y(t)$, $\overline{y}(t)$, $q(t)$, $\overline{q}(t)$, $v(t)$, $\overline{v}(t)$, $w(t)$, $z(t)$. From now on, we assume that all the transfer functions of the systems are with zero initial conditions, unless otherwise specified.

We first provide expressions for the Laplace transforms of the plant input $\overline{u}(t)$ and the plant output $\overline{y}(t)$, given reference $r(t)$ and under injection attacks $w(t)$ and $z(t)$.

PROPOSITION 3.1. *Consider the SISO feedback system with two-way coding under injection attacks depicted in*

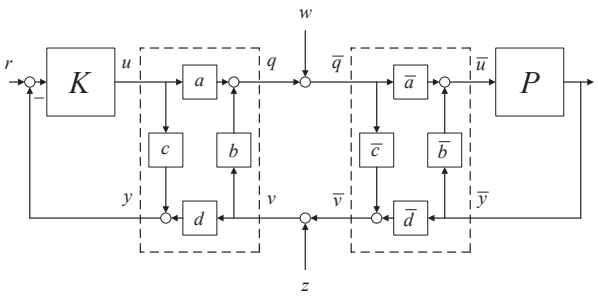

Figure 3: A feedback system with two-way coding under injection attacks.

Fig. 3. Assume that controller K and plant P are LTI with transfer functions $K(s)$ and $P(s)$, respectively, and that the closed-loop system is stable. Then,

$$\overline{U}(s) = \frac{K(s)}{1+K(s)P(s)}R(s) + \frac{a^{-1}[1+cK(s)]}{1+K(s)P(s)}W(s) + \frac{a^{-1}[b-(ad-bc)]P(s)}{1+K(s)P(s)}Z(s), \quad (4)$$

and

$$\overline{Y}(s) = \frac{K(s)P(s)}{1+K(s)P(s)}R(s) + \frac{a^{-1}[1+cK(s)]P(s)}{1+K(s)P(s)}W(s) + \frac{a^{-1}[b-(ad-bc)K(s)]P(s)}{1+K(s)P(s)}Z(s), \quad (5)$$

In what follows, we propose the notion of attack decoupling, which features a strong notion of security in the context of cyber-physical systems; in general, however, it is a more broad control-theoretic notion applicable to any (networked) feedback systems.

Definition 3.2. Consider an LTI feedback control system. An attack is said to be decoupled if the transfer function from attack to plant input/output can be made zero, without nullifying the transfer function from reference to plant input/output.

When the attack is decoupled for a certain attack point, it is as if the path from the attack signal to plant input/output signal is cut off, while not cutting off the signal path from the reference to plant input/output. In general, attack decoupling is a "system-theoretic" notion, which is not restricted to dealing with attacks and is more broadly applicable to disturbances and noises. While within the scope of attack analysis, attack decoupling is a strong notion of security, meaning that the attack response will be zero both in transient phase and in steady state for arbitrary injection attacks, regardless of what the attacker knows or does.

As a matter of fact, attack decoupling is closely related to the notion of disturbance decoupling in geometric control [17]. More specifically, disturbance decoupling

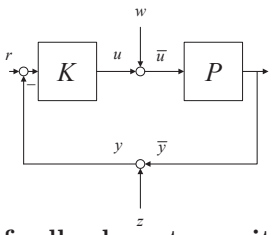

Figure 4: A feedback system without coding.

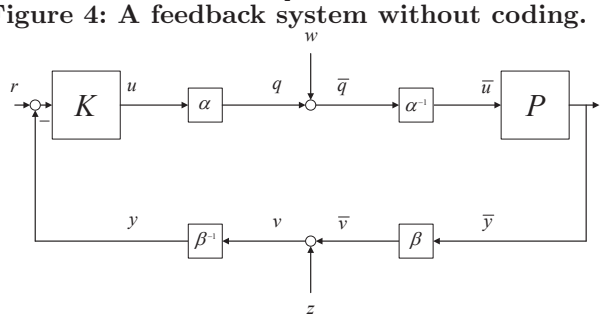

Figure 5: A feedback system with one-way coding.

only requires that the transfer function from the disturbance to plant output to be zero, without requiring the transfer function from the disturbance to plant input to be zero. In this sense, attack decoupling implies and provides a new way to achieve disturbance decoupling, while bringing new perspectives to other relevant topics in geometric control as well.

For conventional feedback control systems without coding, as depicted in Fig. 4, neither attack $w(t)$ nor attack $z(t)$ can be decoupled.

THEOREM 3.3. *Consider the SISO feedback system depicted in Fig. 4. Then, neither attack $w(t)$ nor attack $z(t)$ can be decoupled.*

In the system in Fig. 5 with one-way coding, neither attack $w(t)$ nor attack $z(t)$ can be decoupled.

THEOREM 3.4. *Consider the SISO feedback system depicted in Fig. 5. Then, neither attack $w(t)$ nor attack $z(t)$ can be decoupled.*

For the system shown in Fig. 3 with two-way coding, attack $w(t)$ can be decoupled, and attack $z(t)$ can be decoupled as well.

THEOREM 3.5. *Consider the SISO feedback system depicted in Fig. 3. Suppose that plant $P(s)$ is stabilizable by static output feedback, and that controller $K(s)$ is chosen among such static output-feedback stabilizing controllers, i.e., $K(s) = K \in \mathbb{R}$.*

- *If $c = -1/K$, then attack $w(t)$ is decoupled;*
- *If $b = (ad-bc)K$, then attack $z(t)$ is decoupled.*

Intuitively thinking, in feedback systems without coding as well as systems with one-way coding, there is only

one feedback loop; as such, if the path from the attack signal to plant input/output signal is to be cut off, then the signal path from the reference to plant input/output will inevitably also be cut off. On the other hand, the presence of two-way coding brings additional feedback loops into a feedback system, enabling, probably in a subtle way, the cutting off of the path from the attack signal to plant input/output signal without cutting off that from the the reference to plant input/output.

Note that attack decoupling of $w(t)$ or $z(t)$ requires the co-design of the controller and two-way coding, as well as the sacrifice of control performance since controllers are limited to be static output-feedback.

In the subsequent theorem, it will be shown that for attacks injecting attack signals $w(t)$ and $z(t)$ at the same time (e.g., covert attacks [14]), the attacks $w(t)$ and $z(t)$ cannot be decoupled simultaneously, and hence the attack effect cannot be made completely zero for arbitrary attacks.

THEOREM 3.6. *Consider the SISO feedback system depicted in Fig. 3. Suppose that plant $P(s)$ is stabilizable by static output feedback, and that controller $K(s)$ is chosen among such static output-feedback stabilizing controllers, i.e., $K(s) = K \in \mathbb{R}$. Then, attack $w(t)$ and attack $z(t)$ cannot be decoupled simultaneously.*

This "impossibility theorem" characterizes on a fundamental level why "double-point" attacks are in general more difficult to defend against than "single-point" attacks. We will, however, leave the discussions on the defense against such double-point attacks to future research.

4 CONCLUSIONS

We have introduced the method of two-way coding into feedback control systems under injection attacks. We have proposed the notion of attack decoupling, and it was seen that the controller and two-way coding can be co-designed to nullify the transfer function from attack to plant, zeroing the attack effect both in transient phase and in steady state.

REFERENCES

[1] Edward C. Van der Meulen. 1977. A survey of multi-way channels in information theory: 1961-1976. *IEEE Transactions on Information Theory* 23, 1 (1977), 1–37.

[2] Song Fang, Karl H. Johansson, Mikael Skoglund, Henrik Sandberg, and Hideaki Ishii. 2019. Two-way coding in control systems under injection attacks: From attack detection to attack correction. arXiv:1901.05420.

[3] Riccardo M.G. Ferrari and André M.H. Teixeira. 2017. Detection and isolation of replay attacks through sensor watermarking. *IFAC-PapersOnLine* 50, 1 (2017), 7363–7368.

[4] Riccardo M.G. Ferrari and André M.H. Teixeira. 2017. Detection and isolation of routing attacks through sensor watermarking. In *Proceedings of the American Control Conference*. 5436–5442.

[5] Andreas Hoehn and Ping Zhang. 2016. Detection of covert attacks and zero dynamics attacks in cyber-physical systems. In *Proceedings of the American Control Conference*. 302–307.

[6] Peter F. Hokayem and Mark W. Spong. 2006. Bilateral teleoperation: An historical survey. *Automatica* 42, 12 (2006), 2035–2057.

[7] Karl H. Johansson, George J. Pappas, Paulo Tabuada, and Claire J. Tomlin. 2014. Guest editorial special issue on control of cyber-physical systems. *IEEE Trans. Automat. Control* 59, 12 (2014), 3120–3121.

[8] Fei Miao, Quanyan Zhu, Miroslav Pajic, and George J. Pappas. 2017. Coding schemes for securing cyber-physical systems against stealthy data injection attacks. *IEEE Transactions on Control of Network Systems* 4, 1 (2017), 106–117.

[9] Yilin Mo, Sean Weerakkody, and Bruno Sinopoli. 2015. Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Systems Magazine* 35, 1 (2015), 93–109.

[10] Fabio Pasqualetti, Florian Dorfler, and Francesco Bullo. 2015. Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems. *IEEE Control Systems Magazine* 35, 1 (2015), 110–127.

[11] Radha Poovendran, Krishna Sampigethaya, Sandeep Kumar S. Gupta, Insup Lee, K. Venkatesh Prasad, David Corman, and James L. Paunicka. 2012. Special issue on cyber-physical systems [scanning the issue]. *Proc. IEEE* 100, 1 (2012), 6–12.

[12] Henrik Sandberg, Saurabh Amin, and Karl H. Johansson. 2015. Cyberphysical security in networked control systems: An introduction to the issue. *IEEE Control Systems Magazine* 35, 1 (2015), 20–23.

[13] Claude E. Shannon. 1961. Two-way communication channels. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*.

[14] Roy S. Smith. 2015. Covert misappropriation of networked control systems: Presenting a feedback structure. *IEEE Control Systems Magazine* 35, 1 (2015), 82–92.

[15] Andre Teixeira, Kin C. Sou, Henrik Sandberg, and Karl H. Johansson. 2015. Secure control systems: A quantitative risk management approach. *IEEE Control Systems Magazine* 35, 1 (2015), 24–45.

[16] Anastasios Tsiamis, Konstantinos Gatsis, and George J. Pappas. 2017. State estimation codes for perfect secrecy. In *Proceedings of the IEEE Conference on Decision and Control*. 176–181.

[17] W. Murray Wonham. 1985. *Linear Multivariable Control: A Geometric Approach*. Springer.

[18] Zhiheng Xu and Quanyan Zhu. 2015. Secure and resilient control design for cloud enabled networked control systems. In *Proceedings of the First ACM Workshop on Cyber-Physical Systems-Security and/or PrivaCy*. 31–42.

[19] Quanyan Zhu and Tamer Basar. 2015. Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: Games-in-games principle for optimal cross-layer resilient control systems. *IEEE Control Systems Magazine* 35, 1 (2015), 46–65.