A Short Review of Zero Dynamics Attacks and Mitigation Strategies

Mengxiang Jiang
Department of EECS
Texas A&M University-Kingsville
Kingsville, USA
mengxiang.jiang@students.tamuk.edu

Rajab Challoo
Department of EECS
Texas A&M University-Kingsville
Kingsville, USA
rajab.challoo@tamuk.edu

***Abstract*—A zero dynamics attack is particularly dangerous for cyber-physical systems (CPSs) due to its covert nature and undetectability. However, if the affected continuous-time physical system is of minimum phase, this attack's impact is minimal, even if it goes unnoticed. When digital control is applied, sampling of sensor measurements and using a zero-order-hold for actuation can introduce 'sampling zeros'. These unstable zeros (regardless of the continuous-time system being minimum phase), make the CPS susceptible to a 'sampling zero dynamics attack.' This paper reviews these zero dynamics attacks through some examples and then discusses several strategies to mitigate the impact on system.**

I. INTRODUCTION

In recent years, there has been a swift advancement in computational resources and cyber elements, including networks and digital processes. This progress has allowed traditional control systems to be operated remotely by decoupling the controller from the physical plant, leading to the emergence of cyber-physical systems (CPSs). While this evolution enhances computational efficiency and system flexibility, it also introduces vulnerabilities to the control systems, making them prone to dysfunctional behaviors or deliberate cyber attacks. These vulnerabilities can lead to significant social costs or even loss of human life. The severity of such threats is highlighted by recent incidents like the attack on the SCADA system of a Ukrainian power plant [1].

As a defensive strategy, industrial control systems are commonly equipped with anomaly (or fault) detectors to safeguard against malfunctions and cyber attacks. These detectors are typically integrated with the controller, functioning as a surveillance tool to spot any unusual activities in the plant. To identify faults or attacks, most of these detectors employ a residual generation method [2]. In this approach, a residual is created as the discrepancy between the estimated output and the actual measurement. When the system output strays from its normal behavior, the residual exceeds a pre-set threshold, triggering an alarm. This type of anomaly detection proves effective for certain types of cyber-physical systems that are affected by faults or basic intrusion attempts.

Despite the implementation of anomaly detectors, many CPSs are still vulnerable due to various potent attacks, notably the zero dynamics attack [3]. This type of attack is particularly formidable because it can potentially compromise the system without detection, especially if the attacker has complete knowledge of the system model and it possesses a nonminimum phase zero. Moreover, recent developments in the robust zero dynamics attack have reduced the need for full model knowledge, making this threat even more concerning [4]. Consequently, as long as a system has unstable zeros, it remains susceptible to the zero dynamics attack.

Most CPSs are composed of a continuous time physical plant and a discrete time digital controller, making them sampled data systems. It's important to recognize that even if the continuous time plant is minimum phase, the process of sampling in the data domain can introduce unstable sampling zeros[5]. This is particularly likely when the continuous time plant has a relative degree greater than two and the sampling period is short. Many vital infrastructures, such as nuclear power plants, smart grids, and SCADA systems, tend to be high-dimensional and have large relative degrees, making them susceptible to this 'sampling zero dynamics attack.'

II. Zero Dynamics Attack

Nowadays, the majority of control systems consist of a plant, typically modeled as a continuous time dynamic system, and a digitally-implemented controller, often running on a micro-controller. In this arrangement, the plant's output is measured at

discrete intervals, specifically at each sampling time. This measured output is then utilized to decide the control input for the subsequent sampling interval. Within this framework, the sampled output signal becomes the sole source of information for monitoring the plant's performance.

In this scenario, it's presumed that the network has been breached, leading to a situation where the signal reaching the actuator is a combination of both the control signal and the attack signal. Additionally, this section operates under the assumption that the actuator is outfitted with a zero order hold (ZOH), which is utilized to create the control input that is then applied to the continuous time plant.

For a system that is nonminimum phase, it's possible to engineer a zero dynamics attack that causes the system's state to become unbounded. Concurrently, this attack keeps the output or residual small, ensuring that the attack remains undetected.

As mentioned earlier, a sampled data system using ZOH may become nonminimum phase, even if its continuous-time equivalent is minimum phase. Specifically, when such a system has a relative degree greater than two and employs ZOH with a sufficiently small sampling period, it inevitably transforms into a nonminimum phase system due to the emergence of additional sampling zeros. This suggests that control systems incorporating digital controllers could be more susceptible to zero dynamics attacks compared to systems that rely on continuous-time controllers.

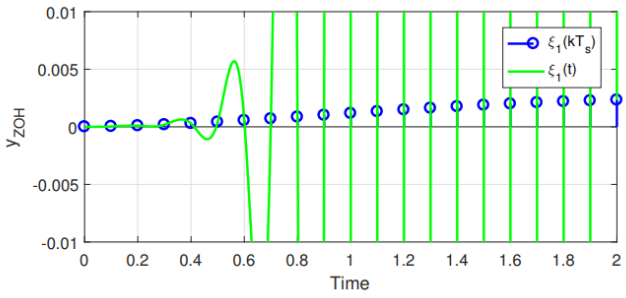


Fig. 1: Sampled output with ZOH under zero dynamics attack[6]

Fig. 1 illustrates the sampled output measurements of a system experiencing a zero dynamics attack. Despite the divergence of both the state variables and the overall system output, the sampled output remains nearly zero. This indicates that detecting a zero dynamics attack in such a scenario is extremely challenging.

III. Mitigation Strategies

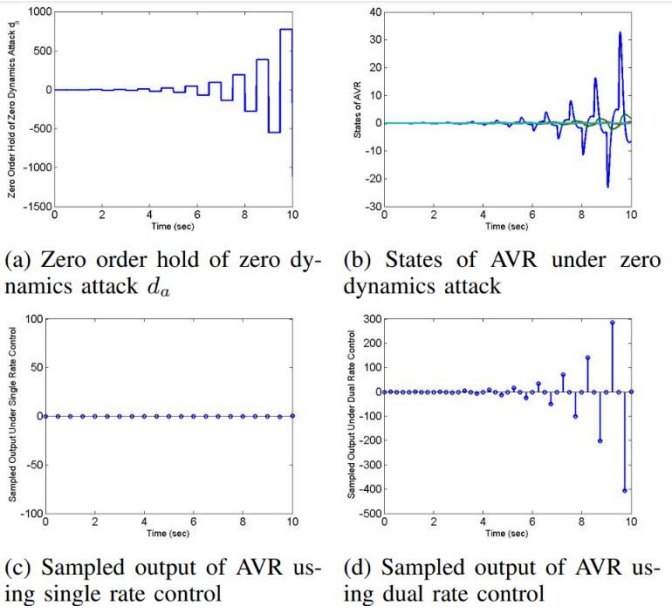
The zero dynamics attack is difficult to detect from the sampled output rendering traditional fault detection methods, which typically rely on output

measurement data, ineffective against this elusive threat. Consequently, alternative solutions have been proposed in the literature, specifically targeting the characteristics of the zero dynamics attack. These remedies, though not numerous, are explored in this section.

One potential strategy involves altering the plant's structure, either by switching out actuators and sensors [3] or by installing an additional modulation block ahead of the control input [7]. After such modifications, the plant's dynamics change into a form unknown to the attacker. This alteration causes the zeros of the modulated plant to shift from their original positions to new locations, unknown to the adversary, thereby hindering the feasibility of a zero dynamics attack. However, even with these changes, the modified plant may still be at risk from stealthy attackers if the relative degree remains unchanged. Recent studies indicate that attackers, even with limited knowledge of the model, can overcome this by accessing disclosure resources (like measurement outputs and control inputs) and employing robust control strategies in their attack plans, known as robust zero dynamics attacks[4]. Furthermore, the positions of sampling zeros can be approximated based solely on the plant's relative degree, especially when the sampling period is very small, regardless of model uncertainty.

Another proposed solution involves using a multi-rate sampling technique, as discussed in [8]. The core concept is that if a zero dynamics attack is designed in the discrete time domain, the continuous time output under the attack might show abnormal behavior, even though it may not be detectable in the sampled measurements. Therefore, if the sampler records the continuous time output more frequently than what the attacker anticipates, the zero dynamics attack could become exposed in the more detailed measurements (See Fig. 2). However, this multi-rate sampling approach might not be as effective if the zero dynamics attack is focused on the intrinsic zeros (i.e., the discrete time zeros corresponding to the continuous time zeros of the plant), particularly when the sampling period is very short. Essentially, a discrete time zero dynamics attack on an intrinsic zero can be seen as a close approximation of its ideally stealthy continuous time counterpart. This approximation becomes increasingly accurate as the

frequency of the holding action's sampling increases. It implies that with a sufficiently small sampling period, the difference between the continuous time outputs under the continuous and discrete time attacks is minimal, making the discrete time zero dynamics attack practically undetectable for a certain period.



(a) Zero order hold of zero dynamics attack d_a

(b) States of AVR under zero dynamics attack

(c) Sampled output of AVR using single rate control

(d) Sampled output of AVR using dual rate control

Fig. 2 Zero dynamics attack on a sampled AVR system [8]

The last potential strategy suggested involve using a generalized hold (GH) over the ZOH [6]. Inspired by the theoretical undetectability of the zero dynamics attack, one approach to counter this type of attack involves relocating the zeros of the sampled data system to a position within the unit circle on the complex plane. By doing so, the impact of the attack could be significantly reduced, making it less damaging. Fig. 3 shows the effect of using GH over ZOH. Even in the scenario where attackers become aware of the newly shifted zeros, their zero dynamics attack would no longer be effective, as all the new zeros are now stable, and all states trajectories remain around and converge to zero, thus mitigating the attack.

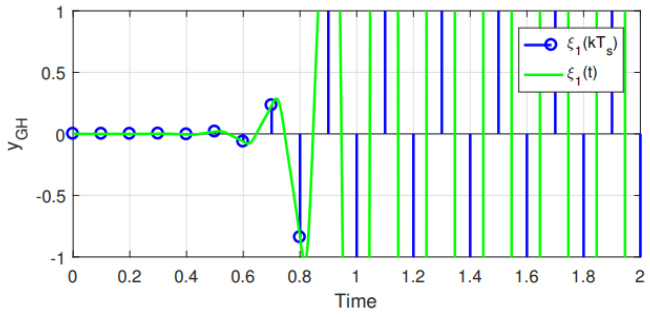


Fig. 3 Sampled output with GH under zero dynamics attack[6]

IV. Conclusion

In conclusion, zero dynamics attacks present a significant threat to CPSs, exploiting vulnerabilities in nonminimum phase systems and remaining undetectable through conventional fault detection methods. The severity of these attacks is amplified by their ability to remain covert, causing potentially catastrophic system failures without early warning. However, the strategies mentioned help mitigate the

impact of such attacks. Each of these methods contributes to enhancing system resilience, highlighting the ongoing evolution in defensive strategies against sophisticated cyber threats. The continuous refinement and implementation of these strategies as well as discovery of new methods are crucial for safeguarding crucial infrastructure and maintaining the integrity of modern control systems.

REFERENCES

[1] R. Langner, "Analysis of the cyber attack on the ukrainian power grid," Joint work between SANS ICS and the Electricity Information Sharing and Analysis Center, 2016.

[2] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," IEEE Transaction on Control Systems Technology, vol. 18, no. 3, pp. 636–653, 2010.

[3] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Revealing stealthy attacks in control systems," in Proc. of 50th Annual Allerton Conference on Communication, Control, and Computing, 2012, pp. 1806–1813.

[4] G. Park, H. Shim, C. Lee, Y. Eun, and K. H. Johansson, "When adversary encounters uncertain cyber-physical systems: Robust zero dynamics attack with disclosure resources," in Proc. of IEEE Conference on Decision and Control, 2016, pp. 5085–5090.

[5] J. I. Yuz and G. C. Goodwin, Sampled-Data Models for Linear and Nonlinear Systems. Springer-Verlag, 2014.

[6] J. Back, J. Kim, C. Lee, G. Park, and H. Shim, "Enhancement of security against zero dynamics attack via generalized hold," in Proc. IEEE 56th Annu. Conf. Decis. Control, 2017, pp. 1350–1355.

[7] A. Hoehn and P. Zhang, "Detection of covert attacks and zero dynamics attacks in cyber-physical systems," in Proc. of IEEE American Control Conference, 2016, pp. 302–307.

[8] M. Naghnaeian, N. Hirzallah, and P. G. Voulgaris, "Dual rate control for security in cyber-physical systems," in Proc. of IEEE Conference on Decision and Control, 2015, pp. 1415–1420.