Application of Failure Detection Methods to Detect Information Attacks on the Control System

Iureva Radda (raddayurieva@gmail.com)
ITMO University https://orcid.org/0000-0002-8006-0980
Margun Alexey
ITMO University

Research

Keywords: Application, detection methods, Information attacks, control system, influence, algorithms

Posted Date: November 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-882662/v1

Eur. Phys. J. C manuscript No.
(will be inserted by the editor)

Application of failure detection methods to detect information attacks on the control system*

Iureva Radda[a,1], **Margun Alexey**[b,1]

[1] ITMO University, Saint Petersburg, Russian Federation

Received: date / Accepted: date

Abstract The paper examines the influence of information attacks on the dynamics of automatic control systems. Comparison of abnormal dynamics of control objects during attacks and device failures is carried out. The similarity of the consequences of information attacks and failures of the control system is analyzed, a method for identifying attacks based on the methods developed for detecting failures is developed. Computer modeling of the influence of information attacks and failures on the control system of a DC motor has been carried out. The simulation results allow making a conclusion about the applicability of the failure detection algorithms for detecting attacks.

1 Introduction

The problem of ensuring the safety of the functioning of control systems is to exclude the influence of failures and attacks on control objects and the environment, i.e., in the elimination of critical failures. This study aims to analyze the similarities between the consequences of attacks on complex technical systems and failures of these systems.

From a safety standpoint, both in the event of a destructive information impact and in case of failure of a system component, its operation should be stopped while maintaining stability, or the performance should be reduced to the prescribed limits if a dangerous failure is inadmissible. To do this, it is necessary to timely detect and isolate information attacks and failures. By isolation, we mean the definition of the type of failure and its localization. Hence the need to create technologies for reliable isolation of information attacks and failures in complex technical systems. In the theory of automatic control, several approaches have been developed that provide detection and isolation of failures based on available measurements and description of the dynamics of control objects. However, no such conceptual schemes, applicable to a wide range of systems, have been proposed for identifying information attacks. In this study, a hypothesis is put forward about the similarity of the influence of failures and attacks on the components of a technical system, which in the future will make it possible to develop approaches for detecting and isolating attacks based on the theory of reliability and the scientific and methodological apparatus of the theory of automatic control.

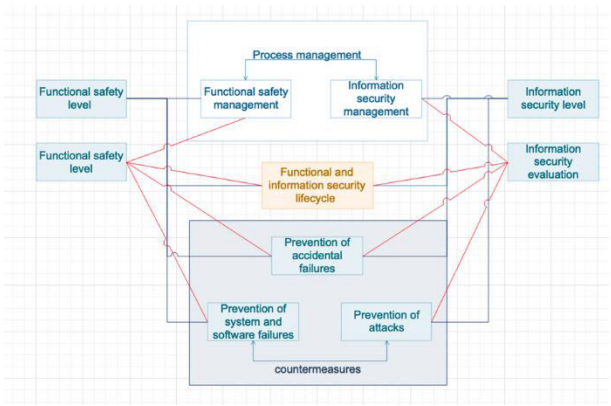

Fig. 1 The structure of requirements for functional and information security of technical systems

The main triad of information security (IS) is availability, integrity, and confidentiality of the data of the technical system, the property of functional security (FS) is to ensure the correct execution of system functions, and in the event of failures, to transfer the control object to a safe state. The analysis of the properties of IS and FS in the complex (figure 1) has become relevant in connection with the development of cyber-physical systems that interact with objects of the real world using global networks and cloud services. These systems can be vulnerable both in the real (physical) world and at the information level, and these levels are inextricably linked in their architecture.

*Thanks to the title
[a]e-mail: raddayurieva@gmail.com
[b]e-mail: aamargun@itmo.ru

Thus, it is required in modern technical systems to ensure reliability in relation to both information attacks and failures. The solution of this problem can be developing a methodology for the simultaneous detection of attacks and failures and their isolation - it is required to determine which failure or attack occurs in the system to take the most effective measures to compensate for the negative effects.

There are several approaches for fault detection: parity relationships, observer based and identification methods [1].

Parity relationship approaches are based on hardware or temporal redundancy. Hardware redundancy solutions require duplication of sensors and actuators. this leads to additional technological and financial costs. The approach based on time redundancy proposes to analyze not the current mismatch of the sensor data with the expected ones, but mismatches at the preceding current moment certain time interval.

Observer-based approaches propose to analyze the residual signal. The residual signal is a mismatch between sensor data and estimates of plant state variables obtained by observers. The problem of fault isolation is solved on the basis of structured residual sets, directional residual vectors and special residual signal generators [2] or filters that sensitive only for special residual signals corresponding to respective faults. Observer-based methods are effective for sensor and actuator faults detection.

Identification-based approaches are used for component fault detection and isolation, where a component fault is a deviation of physical parameters from their nominal value [3].

Hamid Behzad, Alessandro Casavola, Francesco Tedesco and Mohammad Ali Sadrnia [4] propose two fail-safe methods for calculating reconciliation sensors for installations with increased sensitivity, as well as two methodologies for designing a failure detection system. The first is based on a formulation with a linear parameter change, and the second is based on the linear fractional transformation paradigm.

Various approaches are used to detect information attacks—an overview of control systems attacks [5] presents four types of attacks (response and measurement injection, command injection and denial of service) and analyzes the consequences of these attacks on the nodes of the control system. The paper [6] presents an analysis of vulnerabilities and detection of attacks, which was carried out on programmable logic controllers (PLCs), as one of the most important components, on the test bench, and a set of rules was created to detect active start/stop attacks. The analysis used a mirroring method to prevent the detection system from placing additional stress on the existing system and adversely affecting system performance.

In [7] the authors have proposed an algorithm for detecting and preventing DDoS attacks based on network changes and is used to overcome the problem of DDoS attacks and protect routing tasks. Through various transactions, a fault can be identified in each sensor node. The DDoS attack identifier is decoupled from network failures based on the error value.

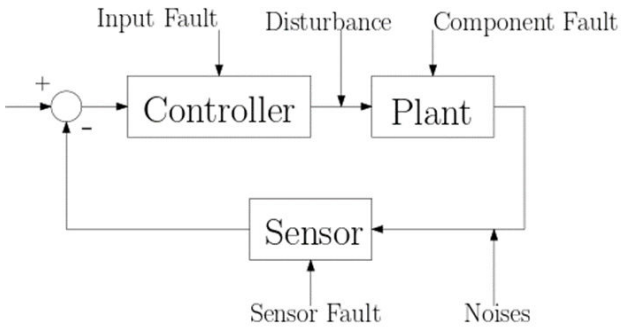

Fig. 2 Typical structure of an automatic control system and possible destructive factors

In [8], a mathematical basis for monitoring attacks on cyber-physical systems is proposed, and a description of the fundamental limitations of monitoring from a system-theoretical point of view and a graph theory point of view.

In [9] the authors proposed methods and measures to counter cybersecurity threats in various approaches to the system design.

Since both failures and information attacks can lead to dangerous consequences for the control system, it is relevant to study the intersection of the information security area and failure detection. The research is based on the hypothesis about the similarity of failures and information attacks on a complex technical system. Both information attacks and failures cause anomalous dynamics of the control object. Analysis of the deviation of the control object's dynamics from the normal mode of operation makes it possible to detect and isolate information attacks and failures.

The contributions of this paper are as follows. Section 2 provides a classification of attacks and failures based on various criteria. Further, in Section 3, portraits of various attacks and failures, their impact on the system dynamics are described. Section 4 provides computer simulations of various failures and attacks on the DC motor control system.

2 Classification of failures and attacks

By cybersecurity, we mean ensuring information and functional security of a cyber-physical system's functioning - a technical system that includes a physical component and a virtual one (algorithms, calculations, channel data transmission medium). The main threats to cybersecurity breaches in complex technical systems are (figure 2):

- information attacks;
- failures in the operation of system nodes, including hardware and software failures and errors.

Considering various classification signs of failures and attacks according to their influence on the dynamics of the

control system from the nature point of view of changes in the parameters that determine the technical state of the object, it is possible to distinguish abrupt or gradual deviations. An abrupt change in parameters can be caused by a critical defect that changes the system's structure (for example, a breakdown of a mechanical part or a failure of a power source) or an information attack (for example, an attacker substitutes the values measured by sensors). A gradual change in parameters is typical for equipment deterioration and change in its parameters due to operating conditions. An example would be the change in resistance of a heating element due to thermal expansion, which affects heating. Similar features can be used to isolate failures and attacks.

From the point of view of the interrelation of destructive processes, they can be independent (only one element of the system works incorrectly, which does not disable others) and dependent (the failure of one element entails several others' failure and the system as a whole). From the functional safety point of view, the latter case is the most critical.

The reasons for equipment failures can be classified: structural, production, operational (Table 1). According to this classification, compliance with the attribute will not be essential for the dynamics of the system, which reduces its value for the task set in the study.

An idle state can persist for a long time, be short-term, or occur periodically under similar conditions. The first case is typical for both attacks and failures. If, after taking measures to prevent the attack, the inoperable state persists, then its reason is the failure, and vice versa. The second case is more typical for information attacks, the third - for technical failures caused by a certain mode of operation of the control object.

For the reliable functioning of control systems, it is necessary to ensure continuous monitoring of its state, a method for detecting failures and attacks, as well as a set of measures to compensate and prevent their influence. This approach will provide a guaranteed level of cyber security in case of hardware and software failures and destructive influences by integrating the scientific and methodological apparatus for identifying and isolating failures for similar tasks when attacking complex technical systems.

3 Analysis of the impact of attacks and failures on the dynamics of control systems

Let us analyze the impact of information attacks on the dynamics of control systems and represent how the anomalous dynamics are interpreted from the fault detection and isolation algorithms' perspective (Table 2). Consider an attacker acting on a controller, plant, and sensors as a source of information attacks.

An attacker can remotely penetrate a controller (a device that calculates control signals and implements control laws). In this case, he can restart the controller or stop its work. During the restart, the controller values are reset (for example, the outputs of the control law's integrators are reset to zero), and the controller will be stopped for a while (initialization). When the controller is stopped, the control signal becomes a constant (including zero). From the theory of automatic control point of view, plant dynamics look like an input disturbance inverse to the control signal. These destructive effects can be identified by actuator fault detection and isolation methods. The following signs can also detect this type of attack: there are no controller output signals (signal is equal to zero or last value of controller), the control signal becomes constant when the sensors data changes.

In the local controller penetration, the attacker intercepts control, i.e., the control signal becomes independent from the sensor data. An approach similar to the described above can be used to detect this type of attack. Possible signs of controller local penetration include a jump-like change of the control signal; the controller output does not correspond to the value calculated on the base of the input data and signals from the sensors; the dynamics of the plant does not correspond to the controller input signal (tracking or stabilization error).

In the case of a local sensor penetration, an attacker alters its measurements. From the automatic control point of view, it is a noise in the measurement channel. However, the amplitude of the noise can reach the limits of the sensor measurement range. Algorithms of sensor faults detection and isolation can be used to detect this type of attack. The following signs correspond to this attack: a rapid increase of the noise amplitude; an abrupt change of the sensor signal; the signal measured by the sensor does not correspond to the predicted value calculated based on the control signal and the plant model.

The controller stops its operation when a remote denial of service attack is implemented. The same features characterize this attack as remote penetration of the controller.

In remote denial of service, signals from the sensor are stopped (nothing, zero, or the last sent value comes as a measurement signal. The same approaches and features for local sensor penetration can be applied to detect this type of attack.

Table 1: Comparison of attacks from the point of view of theory of automated control and IS

Classification feature	Values (nature of change) of the classification feature	Type of failure	Reasons
The nature of the change in the parameters that determine the technical state of the object	Abrupt change in one or more parameters	Sudden	Internal defects, operator errors, operational disturbances, local penetration
	Gradual change in one or more parameters	Gradual	Aging of materials, corrosion, wear of parts, etc.
Interrelation of failures	The failure of an element is not caused by damage or failure of other elements of this object.	Independent	–
	Element failure due to damage or failure of other elements	Dependent	Damage and failures of other elements of an object or system.
Origin of failure	Violation of established rules and (or) design standards, imperfection of accepted design methods.	Structural	Errors in the development and design of an object, underestimation of safety margins, violation of GOST standards, etc.
	Violation of the established process of manufacturing or repairing an object, imperfection of manufacturing technology	Manufacturing	Failure to comply with documentation standards, use of low-quality materials and components, insufficient level of production quality control, etc.
	Violation of the established rules and (or) operating conditions	Operational	Errors of low-qualified service personnel, ignoring / violation of the rules of technical documentation, as well as aging and wear of equipment for the above reasons.
Stability of an inoperative state	Stable persistent	Stable	Change of object parameters, irreversible damage to system elements
	It remains for a short time, after which the operability is self-healing or restored by the operator without repairs	Self-eliminating (sporadic failure)	Short-term external influences, short-term change in object parameters
	It has the same character, arises and removes itself many times	Intermittent	External interference and impacts that go beyond the permissible technical limits and are reversible.

Table 2 Comparison of attacks from the point of view of theory of automated control and IS

Attack	Object	Action	According to fault
Remote penetration	Controller	Remote restart or shutdown	Input disturbance equal to inverse of input signal
Local penetration	Controller	Control Intercept	Input disturbance with amplitude up to input signal range
	Sensor	Data modification	Output noises with amplitude up to sensor measurement range
Remote denial of service	Controller	Full stop	Input disturbance equal to inverse of input signal
	Sensor	Full stop	Output noises with amplitude up to sensor measurement range
Jamming and data spoofing on sensor	Sensor	Jamming or data spoofing	Output noises with amplitude up to sensor measurement range.
Decommissioning of component	Component	Destructive effect, manifested in the unstable functioning of the component	Dramatic deviation between measured and predicted dynamics of the plant

During sensor jamming or data substitution, the sensor signal does not correspond to measured physical values. The following signs can accompany this attack: a sharp increase of the noise amplitude, an abrupt change of the sensor signal; the measured output of the plant does not correspond to the predicted one based on the model of the plant. Sensor fault detection and isolation algorithms are effective for the detection of this attack.

In the case of an attack that entails the failure of one of the plant components, its dynamics becomes unpredictable. This attack can be determined by the method of elimination: if all previous faults and attacks are excluded, and the component's dynamics do not correspond to the nominal, then the dynamic model is incorrect. Therefore, one of the components has failed. Also, it is possible to build a set of models with faults. The attack detection can be based on the similarity of the system's behavior and faulty system dynamics.

4 Modeling attacks on a control system

Consider DC motor. Its dynamics is described by equations:

$$L\frac{dI}{dt} + RI = U - E_b,$$
$$L\frac{dI}{dt} + RI = U - E_b,$$
$$J\dot{\omega} = M - M_{fr},$$

where ω is an angular velocity, I is a current, L is an armature inductance, R is an armature resistance, U is an input voltage, $E_b = k_e\omega$ is a back-EMF, k_e is a constant, $J = J_d + J_m$ inertia momentum, J_d is a rotor inertia momentum, J_m is a load inertia momentum, $M = k_m I$ is a motor force momentum, k_m is a constant, $M_{fr} = k_f\omega$ is a friction momentum, k_f is a friction coefficient.

Rewrite model under faults in state space representation:

$$\dot{x} = Ax + Bu + f_a$$

$$y = Cx + f_s$$

Where $x^T = [\omega\ i]$ is a state vector,

$$A = \begin{bmatrix} -k_f/J & k_m/J \\ -k_e/L & -R/L \end{bmatrix} = \begin{bmatrix} a1 & a2 \\ a3 & a4 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 \\ 1/L \end{bmatrix},$$

f_a is an impact of attack on the controller, f_s is an impact of attack on the sensor.

Assume that motor equipped with a velocity sensor. Therefore:

$$C = [1 \quad 0].$$

4.1 Attack on the executive device

Controller that calculates the supply voltage with an integrated or connected driver is an actuator for a DC motor. Figure 3 shows the signal of the DC motor angular velocity sensor during attacks on the controller, where ω is a sensor data without attack, ω_{local} is a sensor signal under controller local penetration, ω_{remote} is a sensor signal under controller remote penetration. Local penetration simulates the case when attacker intercepted control of the motor and applied excess voltage from 5 to 10 seconds. The controller was disconnected from 5 to 10 seconds under remote penetration.

The dynamics of the output in the considered cases is identical to the dynamics in the case of actuator fault. Therefore, abnormal behavior during attacks on the controller can be detected using the state observers-based methods of fault detection. Attack isolation can be implemented using methods of actuator fault isolation, for example, [2].

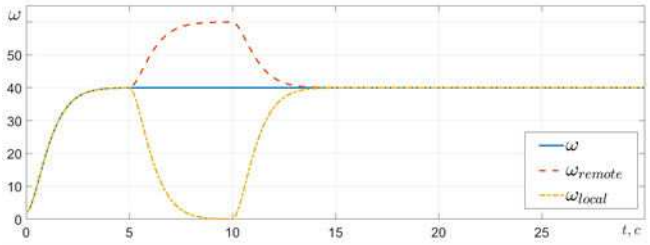

Fig. 3 Readings of the DC motor speed sensor during attacks on the controller

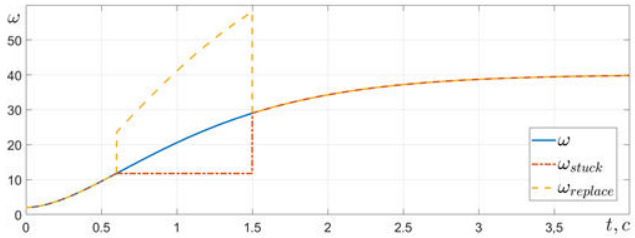

Fig. 4 Readings of the angular velocity sensor when attacking it

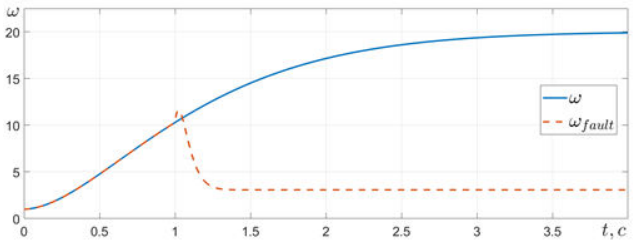

Fig. 5 Readings of the angular velocity sensor when attacking it

4.2 Attack on the sensor

Consider the effect of sensor attacks on motor velocity transients. Figure 4 shows examples of the angular velocity sensor signal during attacks on it. Graph ω illustrate transients without attacks.

Graph ω_{stuck} illustrates jammed sensor data from 1.6 to 1.5 seconds. A sensor jamming can be caused by local penetration or a remote denial of service. Case of sensor data replacement on $\omega_{replace}$ is presented in figure 4.

Simulation results show that sensor signals during information attacks are similar to signals during failures. Thus, we can conclude that the sensor signals are similar in the cases of faults and attacks. Therefore, such methods of sensor fault detection and isolation as based on observers and generators of residuals, hardware and time redundancy can be used to detect attacks on the sensor.

4.3 Component failure/component attack

An attack on one of an automatic control system component can lead to parametric and structural disturbances. Figure 5 graphs of velocity sensor under normal functioning (ω) and under sufficient deviation of motor parameters (ω_{fault}) is presented below.

The similarity of the behavior of the system under component fault and under attack acting on plant parameters is due to the same physical impact on the plant. Therefore, it is advisable to use such methods based on the identification of the parameters of the plant for attack detection and isolation [3] as gradient approach, the least squares and dynamic regressor extension and mixing [10]. It is difficult to directly determine which element has been attacked leading to plant structure changes since the behavior of the system becomes unpredictable. In this case, it is advisable to build faulty plant models under such attacks with further analysis of dynamics similarity.

Conclusion

In the paper is analyzed the impact of information attacks and failures on automatic control systems, considered the reasons and behavioral portraits of various types of attacks on technical systems and failures of actuators, sensors, and components. The analysis revealed that the detection and isolation of failures can be used to detect and isolate a wide class of attacks on controllers, measuring devices and control system components. Computer modeling using a DC motor as an example revealed that the dynamics of control systems subjected to information attacks is similar to the dynamics of control systems with failures. This conclusion can be extended to a wide class of technical systems. Based on the results obtained, in the future, a structure for ensuring information and functional security can be developed (Figure 1), based on the scientific and methodological apparatus for detecting and isolating failures of technical systems, which will increase the level of reliability, timely identify failures and information attacks within one control system, take timely measures to compensate their impact, reduce the time to restore the correct operation of the system.

6 List of Abbrevations

- IS – information security;
- FS – functional security;
- DC motor – direct current motor;
- PLC - programmable logic controllers;
- DDoS attack – distributed denial-of-service attack;
- GOST – set of technical standards maintained by the Euro-Asian Council for Standardization, Metrology and Certification (EASC);
- EMF – electromotive force

7 Declarations

7.1 Availability of supporting data

The data that support the findings of this study are available from the corresponding author, Iureva Radda, upon reasonable request.

7.2 Competing interests

There are no relevant financial or non-financial competing interests to report.

7.3 Funding

Not applicable.

7.4 Authors' contributions

Iureva Radda: Conceptualization, Methodology, Data curation, Writing - Original draft preparation, Validation, Writing-Reviewing and Editing. Margun Alexey: Software. Writing - Original draft preparation, Visualization, Investigation.

7.5 Acknowledgements

Not applicable.

References

1. J. Chen and R. J. Patton, Robust Model-Based Fault Diagnosis for Dynamic Systems (Kluwer Academic Publishers, Boston, MA, U.S.A., 1999)
2. R.J. Patton, J. Chen, Observer-based fault detection and isolation: Robustness and applications, Control Engineering Practice **5**, 5, pp. 671-682 (1997)
3. R. Isermann, Supervision, fault-detection, and fault-diagnosis methods, An introduction, Control Engineering Practice, **5**, 5, pp. 639-652 (1997)
4. Hamid Behzad, Alessandro Casavola, Francesco Tedesco & Mohammad Ali Sadrnia. Fault-Tolerant Sensor Reconciliation Schemes based on Unknown Input Observers (2018). doi.org/10.1080/00207179.2018.1484568
5. Morris, Thomas & Gao, Wei. Classifications of Industrial Control System Cyber Attacks. First International Symposium for ICS & SCADA Cyber Security Research (2013)
6. Ercan Nurcan Yılmaza, Serkan Gönenb. Attack detection/prevention system against cyber-attack in industrial control systems (2018). doi.org/10.1016/j.cose.2018.04.004
7. Rathika, R. K., and A. Marimuthu. An improve attacks in nuclear power plants machine monitoring. Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), pp. 1-7. (2017)
8. Fabio Pasqualetti, Florian Dorfler, and Francesco Bullo. Attack Detection and Identification in Cyber-Physical Systems in IEEE Transactions on Automatic Control, **58**, 11, pp. 2715-2729 (2013). DOI: 10.1109/TAC.2013.2266831.
9. Danenkov I., Kolesnikova D., Babikov A., Iureva R. Security by Design Development Methodology for File Hosting Case. Smart Education and e-Learning **188**, (2020). doi.org/10.1007/978-981-15-5584-8_33
10. Belov A., Aranovskiy S., et al. Enhanced Parameter Convergence for Linear Systems Identification: The DREM Approach. Proceedings of the 2018 European Control Conference (2018).