UDC 004.056.55:623.746.9:681.3.06

Stealthy Cyberattacks on Control Systems Using an Adaptive Soft-Constrained Optimization Method

Oleksii Novikov[1], Mykola Ilin[1], Iryna Stopochkina [1] and Volodymyr Duduladenko[1]

[1] *National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine*

Abstract

This paper presents a novel approach for designing stealthy cyberattacks on automated control systems of critical infrastructure. The core idea lies in employing an adaptive soft-constrained optimization method, which simultaneously maximizes the impact functional of the attacker while keeping the attacked trajectory within the invisibility range of a standard fault detection mechanism. The proposed approach is based on a variational problem formulation, the construction of adjoint equations, and a gradient-based procedure with dynamic penalty parameter updates. Numerical simulation is conducted on a second-order test dynamic system. The results demonstrate the algorithm's effectiveness and convergence, as well as the feasibility of generating a controlled attack that successfully bypasses WLS-based detection methods. The method can be used to test the resilience of industrial systems to cyber threats through security scenario modeling.

Keywords: cyberattacks, critical infrastructure, stealthy attacks, automated control systems, optimization methods

Introduction

In recent years, there has been a dramatic increase in the number of cyberattacks on critical infrastructure. This alarming trend is largely driven by the increasing vulnerability of modern automated control systems of technological processes, such as Industrial Control Systems (ICS), Supervisory Control and Data Acquisition (SCADA), Distributed Control Systems (DCS), and systems based on Programmable Logic Controllers (PLC). These modern control systems, especially those in critical infrastructure, are increasingly integrated into digital networks using open communication protocols, which significantly raises the risk of cyber intrusions.

Lower-level automated control systems, particularly those based on PLCs, remain among the most vulnerable to cyberattacks. At the same time, these systems are responsible for managing key parameters of critical technological processes.

The most common forms of attacks include Denial-of-Service (DoS) attacks (overloading PLCs), replay attacks, false data injection, and more sophisticated approaches such as zero dynamics attacks, covert attacks, and stealthy attacks that are capable of bypassing traditional monitoring mechanisms [1–5]. Among the most dangerous are stealthy attacks, which distort control or measurement signals while remaining undetected by standard detection tools.

The threat level posed by such attacks depends on the effectiveness of detection systems, which include classical diagnostic techniques as well as modern behavior-based, machine learning, and data-driven approaches. Numerous studies [6–12] have analyzed conventional detection methods, including Bayesian hypothesis testing, Kalman filter-based χ^2 detectors, Fault Detection and Isolation (FDI) approaches, and the Weighted Least Squares (WLS) method. WLS minimizes the quadratic functional of measurement residuals, triggering a fault detector if this quantity exceeds a defined threshold. However, the advancement of these techniques introduces new challenges: it becomes necessary to model complex attacks capable of adapting to system parameters while remaining undetectable.

Traditional methods, especially those lacking adaptation to environmental changes, may fail to identify targeted attacks that exploit the structure of the diagnostic system. Under such conditions, the development of specialized cyberattack scenarios that account for stealth constraints becomes an effective strategy for testing cyber defense mechanisms and revealing potential vulnerabilities in existing systems.

To inflict maximum harm, adversaries design various attack scenarios on automated control systems. The goal of such attacks may be to distort the control signal in a way that maximizes the deviation of the system state from its nominal trajectory. In the literature, multiple optimization-based approaches have been explored to model such attacks, including Lagrange multipliers, penalty methods, barrier methods, and others [13–16]. Penalty methods incorporate constraints by adding extra terms to the objective functional but require precise tuning of coefficients. Barrier methods operate within the interior of the feasible set. The Lagrangian method explicitly includes constraints but requires reformulating the problem into a minimization form. Despite superficial similarities, these methods differ in physical meaning, advantages, and drawbacks.

In this work, we implement an adaptive soft-constrained optimization method with a dynamically updated penalty multiplier. This enables adaptive convergence to the detection boundary without destabilizing the system. **Research Objective.** The objective of this work is to develop and analyze the parameters of a stealthy attack on the control system of a critical infrastructure facility, intended as a tool for testing cyber defense mechanisms. The attack aims to maximize the deviation of the system's state trajectory from its nominal behavior while remaining undetected by a standard fault detection mechanism

1. Attack Model on the Control System

Let us consider a dynamic system described by:

$$\frac{dx(t)}{dt} = Ax(t) + Bu(t), \quad (1)$$

$$x(t_0) = x_0, \quad (2)$$

$$y(t) = Cx(t), \quad (3)$$

where $x(t)$ is n - dimensional vector of the state physical system; $y(t)$ is l - dimensional vector of the measurement system sensors; A, B and C – are known matrices of appropriate dimensions; $u(t)$ is k - dimensional the control input vector. Equations (1), (2) and (3) describe the system dynamics and the measurement model, respectively.

Let us consider the problem of optimal control of the system state $x(t)$ governed by (1), (2), using a feedback law and a quadratic cost criterion [13–17]:

$$F = \int_{t_0}^{t_k} [x^T(t)Qx(t) + u^T(t)Ru(t)]\, dt, \quad (4)$$

where Q and R - are known weighting matrices.

The optimal control law for the system (1), (2) can be written as [13]:

$$u(t) = -K(t)x(t), \quad (5)$$
$$K(t) = R^{-1}B^T P(t),$$

where matrix $P(t)$ is the solution to the nonlinear Riccati equation:

$$\frac{dP(t)}{dt} = -P(t)A^T - AP(t) + P(t)BR^{-1}(t)B^T P(t) - Q, \quad (6)$$

$$P(t_k) = P_k. \quad (7)$$

This optimal control framework is widely used in ICS, SCADA, DCS, and PLC-based control systems in critical infrastructure.

Now, consider a cyberattack targeting the automatic optimal control system defined by (1) – (7). The main objective of the attack is to disrupt normal system operation by distorting the control signal to maximize the system state deviation. The attacker is assumed to be capable only of additively modifying the control input components (see Fig. 1).

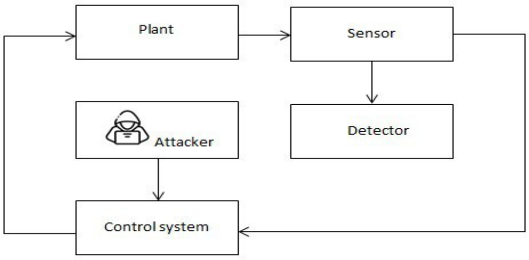

Figure 1: Cyberattack on the control input of an automated control system

A key feature of the scenario under study is the presence of a fault detector, which performs anomaly or intrusion detection. A successful stealthy attack must remain invisible to this detector. In other words, the attacker implements a stealthy attack, carefully designed to evade detection.

To execute such an attack, the adversary obtains access to key system components such as software code, control logic in the PLC, and the fault detection algorithm. Special attention is given to studying the fault detection method in use, enabling the attacker to bypass it or remain undetected. The attacker's influence is limited to crafting a malicious control input.

This leads to the attacked control model:

$$\tilde{u}(t) = u(t) + u_a(t), \tag{8}$$

where $u_a(t) \in U_{\text{доп}}$ – is the stealthy attack signal: malicious data added to the system control input.

The attacker's maximization strategy is defined via an impact functional:

$$\Phi(u_a) = \int_{t_0}^{t_k} x^T(t) W x(t) dt \to \max_{u_a \in U_{per}} \tag{9}$$

where W - is a known weighting matrix.

To ensure the stealthiness of the attack, the impact functional (9) is constrained by a detection threshold, formulated using the Weighted Least Squares (WLS) method [6–12]:

$$J(u_a) = \int_{t_0}^{t_k} [Cx(t) - y(t)]^T S^{-1} [Cx(t) - ytdt \leq \delta, \tag{10}$$

where $y(t)$ – is the measurement vector used by the fault detector, S – is a known weighting matrix, and δ – is the detection threshold.

These stealth conditions (10) are typically enforced by standard fault detectors in automation systems. However, a knowledgeable attacker can exploit this method to design attacks that minimize detection risk. Hence, the system's security depends on the robustness of its architecture, the reliability of fault detection algorithms, and the ability to adapt to potential threats.

2. Optimization Problem for Stealthy Attack Design

We now rewrite the impact functional (9) under attack (8) as:

$$\Phi(u_a) = \int_{t_0}^{t_k} \tilde{x}^T(t) W \tilde{x}(t) dt \to \to \max_{u_a \in U_{per}} \tag{11}$$

where $\tilde{x}(t)$ – evolves under the influence of the malicious input u_a.

The stealthiness condition (10) and the system (1) – (7) under attack (8) can be rewritten as:

$$J(u_a) = \int_{t_0}^{t_k} [C\tilde{x}(t) - -y(t)]^T S^{-1} [C\tilde{x}(t) - y(t)] dt \leq \delta, \tag{12}$$

$$\frac{d\tilde{x}(t)}{dt} = A\tilde{x}(t) + B[-K(t)\tilde{x}(t) + + u_a(t)](t), \tag{13}$$

$$\tilde{x}(0) = \tilde{x}_0, \tag{14}$$

where the control matrix $K(t)$ is given by (5).

The goal is to find a malicious control $u_a(t) \in U_{\text{доп}}$ that maximizes the impact functional (11) while satisfying the stealth condition (12) for the system defined by (13)–(14).

3. Lagrangian Formulation with Adaptive Penalty

To solve the constrained maximization problem, we apply a modified Lagrangian method [17]. Unlike the classical Karush–Kuhn–Tucker (KKT) formulation for constrained minimization, our approach implements a gradient ascent strategy for functional maximization with a dynamically updated penalty term.

We define the augmented Lagrangian functional:

$$L(\tilde{x}(t), u_a, \lambda) = \int_{t_0}^{t_k} \Big\{ \tilde{x}^T(t) W \tilde{x}(t) + \lambda TtAxt-B-Ktxt+uat-dxtdt-\xi Cxt-y tTS-1Cxt-ytdt, \tag{14}$$

where λ(t) - is the Lagrange multiplier (adjoint variable), and parameter ξ is a scalar adaptive penalty multiplier, the constraint on $u_a(t) \in U_{per}$ is enforced in the final stage.

This method falls within the class of soft-constrained adaptive penalty methods, suitable for problems with complex or fuzzy constraints. It avoids strict KKT conditions while enabling gradient ascent to achieve a local maximum of $L(\bullet)$, naturally preserving the attacker's objective of maximizing impact while staying undetected.

Applying the variational principle and setting $\partial L / \partial \tilde{x}$ = 0, we obtain the adjoint equation:

$$\frac{d\lambda(t)}{dt} = 2W\tilde{x}(t) + (A^T - K^T B^T)\lambda(t) - 2\xi C^T S^{-1}[C\tilde{x}(t) - y(t)] \quad (15)$$

$$\lambda(t_k) = 0. \quad (16)$$

The necessary optimality condition with respect to the malicious control $u_a(t) \in U_{per}$ is:

$$\delta L(u_a)\frac{dL}{du_a} = \delta u_a = 0 \quad (17)$$
$$\forall u_a(t) \in U_{per}.$$

We supplement condition (17) for $u_a(t) \notin U_{доп}$:

$$\frac{dL}{du_a} = 0 \ \forall u_a(t) \in U_{per}. \quad (18)$$

Thus, the gradient of (14) with respect to $u_a(t)$, becomes:

$$\frac{dL}{du_a} = \int_{t_0}^{t_k} B^T \lambda(t). \quad (19)$$

The optimal stealthy attack $u_a(t)$, is found via an iterative gradient ascent procedure:

$$u_a^{i+1} = Pr\{u_a^i + \alpha\frac{dL^i}{du_a}\}, \quad (20)$$

where $Pr\{\bullet\}$ – projects the solution $u_a^{i+1}(t)$ onto the feasible control set $u_a(t) \in U_{доп}$, i - is the iteration index, a α is the gradient ascent step size, u_a^0 is the initial malicious control guess.

The penalty multiplier ξ is adaptively updated using:

$$\xi^{i+1} = max\{0, \xi^i + \alpha_\xi(J^i - \delta) - \beta\xi^i\}, \quad (21)$$

where α_ξ - is the update step, β – is the damping coefficient that restrains the growth of ξ.

The procedure terminates once:

$$\mid J^i - J^{i+1} \mid \le \varepsilon, \quad (22)$$

where ε is a predefined accuracy threshold.

4. Algorithm for Stealthy Attack Design via Adaptive Penalty Method

Combining expressions (11) – (22), we outline the full algorithm to compute the stealthy control u_a, that maximizes the impact functional $\Phi(u_a)$ under the stealth constraint $J(u_a) \le \delta$ enforced via an adaptive penalty multiplier ξ:

1. Initialization:

At iteration i = 0, provide initial guess for u_a^0 and ξ^0, set step sizes α, α_ξ - gradient ascent and penalty update ξ, damping coefficient β, and convergence threshold ε.

2. Main Iterative Loop (for *i = 0, 1, 2, ...*):

2.1. Forward Simulation: Integrate the system dynamics (13), (14) using u_a^i to compute state trajectory $\tilde{x}(t)$.

2.2. Functional Evaluation: Compute the impact functional Φ^i and the stealth constraint J^i using (11), (12).

2.3. Adjoint Equation Solution: Integrate the adjoint equation (15), (16) backward in time to obtain $\lambda^i(t)$.

2.4. Gradient Computation: Evaluate the gradient of the Lagrangian dL^i/du_a using (19).

2.5. Update Control Input: Update u_a^i using (17)-(20).

2.6. Updating the adaptive penalty multiplier: We update the multiplier ξ^i according to (21).

2.7. Stopping Criterion: If $\mid J^i - J^{i+1} \mid \le \varepsilon$, terminate. Otherwise, proceed to next iteration.

Control example and initial data. Let us consider a control example of a numerical experiment conducted to verify the performance of the algorithm for constructing a stealth attack using the adaptive soft-constrained optimization method. According to the problem statement, the purpose of the work is to develop and study a method for determining the parameters $u_a(t)$ of a stealthy attack on the control system of a critical infrastructure object with the maximization of the impact functional Φ subject to the constraint on stealth through $J \le \delta$. The optimization procedure is implemented in the form of an adaptive penalty with a variable multiplier ξ. As a test example, a second-order linear system with the parameters given in Table 1 was used. Note that the Riccati function, as well as the column vector of measurements $y(t)$, were calculated separately outside the gradient search procedure.

5. Simulation results

To assess the efficiency of the proposed algorithm, we performed numerical simulations. The results are shown in Figures 2–5.

In Fig. 2. we will see convergence of the impact functional $\Phi(u_a)$ and stealthy $J(u_a)$ in the process of the gradient algorithm. Monotonic saturation of $\Phi(u_a)$ and decrease of $J(u_a)$ with each iteration are observed, which indicates the effectiveness of adaptive updating of the penalty multiplier ξ and achievement of the detection limit $J(u_a) \le \delta$.

Note that the Riccati function, as well as the column vector of measurements $y(t)$, were calculated separately outside the gradient search procedure.

Fig. 3 shows the attack signal u_a, which dynamically evolves at each iteration, aiming to maximize $\Phi(u_a)$ and suppress $J(u_a)$, being in boundaries of permissible values $u_a(t) \in U_{per}$. Convergence to a steady control profile is observed in the final iterations.

In Fig. 4,5 evolution of the adaptive penalty multiplier ξ are shown, and system state trajectories $\tilde{x}(t)$ - with and $x(t)$ - without attack.

Table 1 Initial values for the experiment

Parameter	Designation	Value
Dimensionality of matrices	A, C, K, P, Q, R, S, W	2x2
Dimensionality of column vectors	$x(t), \tilde{x}(t), \lambda(t), y(t), u_a(t)$	2x1
Matrix of model coefficients	A	0; 1; -6; -5
Matrix of controls	B	0; 1; 0; 0
Matrix of measurements	C	1; 0; 0; 1
Matrix of feedback	K	1.5; 0; 0; 1.5
Weight matrices of the criterion of the optimal control system	Q R	1; 0; 0; 1 1; 0; 0; 1
Fault detector weight matrix	S	1.5; 0; 0; 1.5
Impact functional weight matrix	W	5.5; 0; 0; 5.5
Model initial conditions	$x(0)$	0; 1
Adjoint equation final conditions	$\lambda(t_k)$	0; 0
Study period and time step	$t_0, t_k, \Delta t$	0; 4; 0.01
Gradient procedure step size and number	$\alpha, \alpha_\xi, \beta$, δ, N	0.00001; 6.0; 0.035; 0.1; 200
Starting value of attack control and penalty multiplier	u_a^0, ξ^0	0.0; 1.0

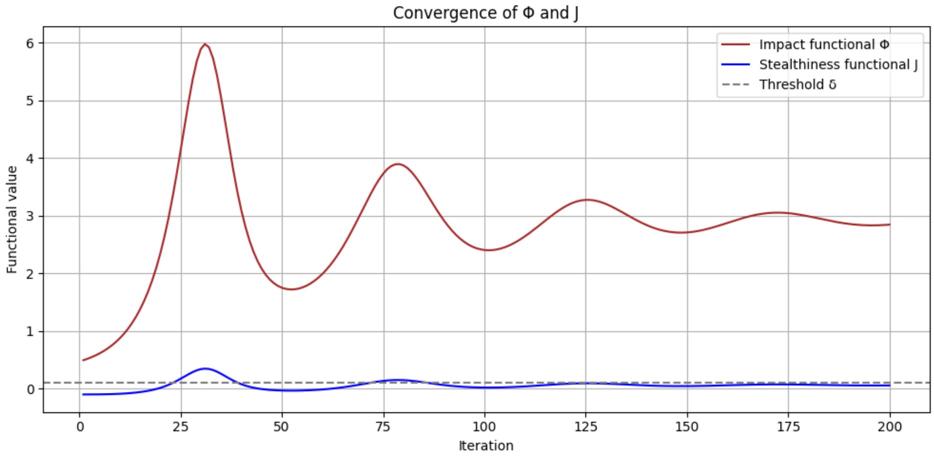

Figure 2: Convergence of the impact functional $\Phi(u_a)$ and stealthy $J(u_a)$ in the process of the gradient algorithm

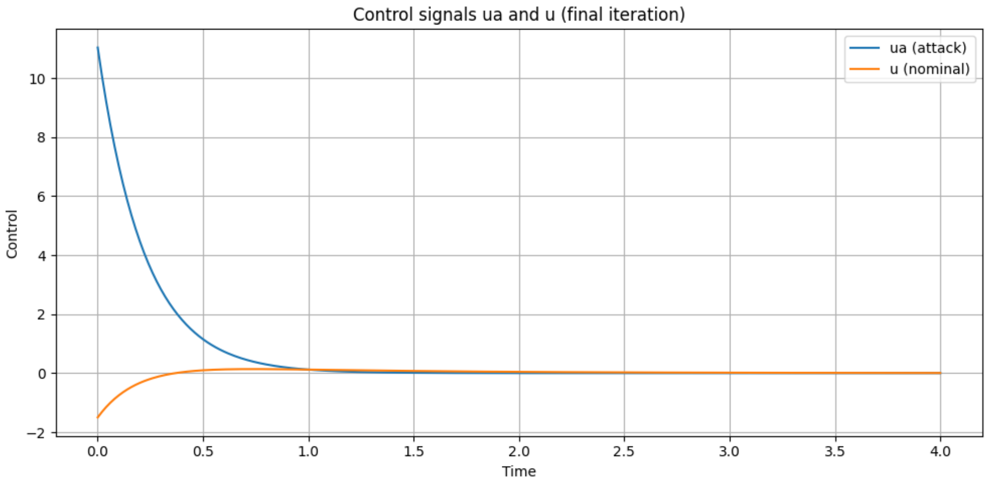

Figure 3: Evolution of the malicious control signal u_a

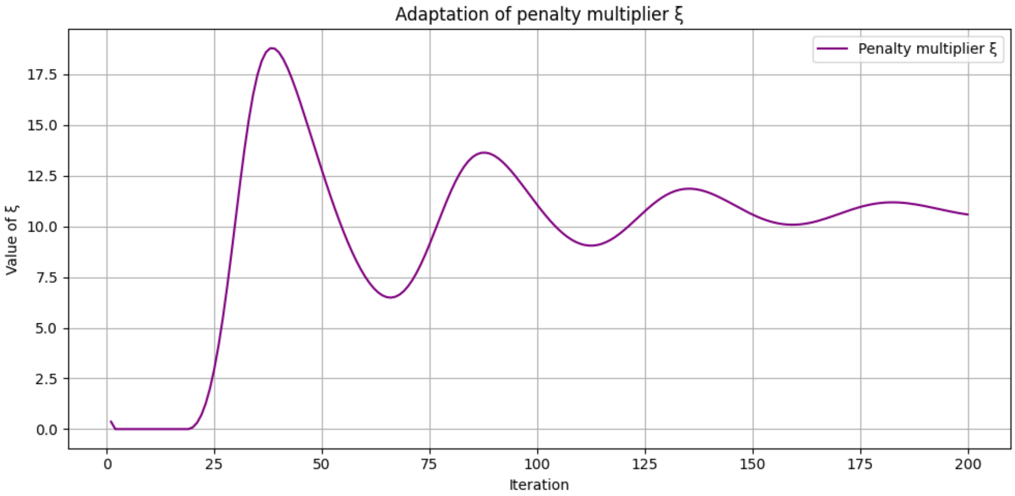

Figure 4: Evolution of the adaptive penalty multiplier ξ

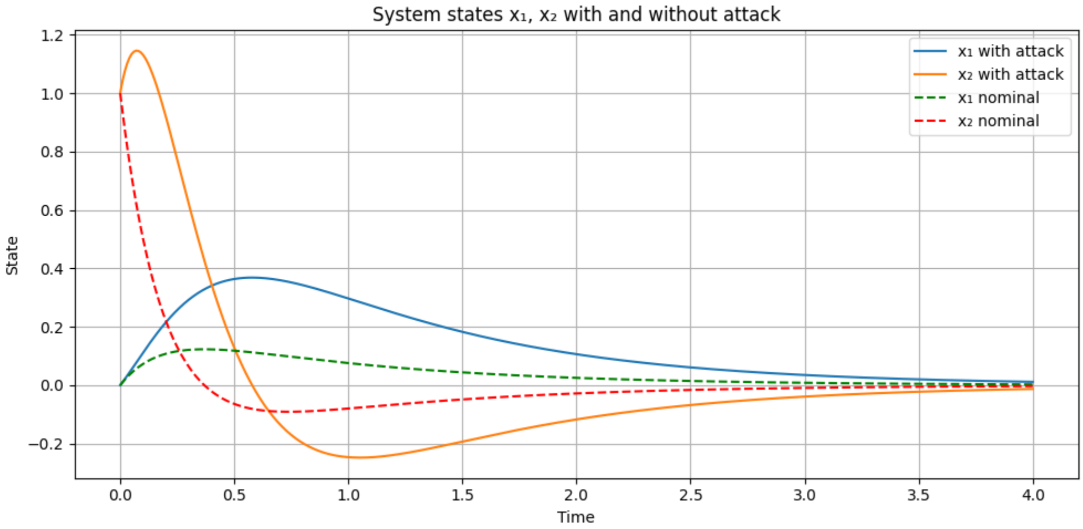

Figure 5: System state trajectories $\tilde{x}(t)$ - with and $x(t)$ - without attack

Analyzing the results of the computational experiment, we can conclude that the iterative gradient procedure for searching for control attack data converges to the solution u_a. The latter provides an opportunity to conclude about the performance of the proposed method and algorithm.

Conclusions

This work implements a novel approach for modeling stealthy cyberattacks based on a soft-constrained optimization strategy using an adaptive penalty method. This approach allows the attacker to enhance impact while regulating detectability through an automatically tuned multiplier. The gradient ascent procedure ensures convergence to a solution that reaches the detection boundary while maximizing disruption.

Simulation results confirm the feasibility of the algorithm and its suitability for generating attack scenarios used in testing the resilience of industrial control systems to cyber threats. The approach can serve as a tool for validating the robustness of real-world detection and diagnostic mechanisms.

References

[1] Teixeira A., Shames I., Sandberg H., Johansson K.H. A secure control framework for resource-limited adversaries. *Automatica*, 2015, vol. 51, pp. 135–148.

[2] Mo Y., Sinopoli B. Secure estimation in the presence of integrity attacks. *IEEE Trans. Automat. Contr.*, 2015, vol. 60, no. 4, pp. 1145–1151.

[3] Pasqualetti F., Dorfler F., Bullo F. Attack detection and identification in cyber-physical systems. *IEEE Trans. Automat. Contr.*, 2013, vol. 58, no. 11, pp. 2715–2729.

[4] Urbina D.I., Giraldo J.A., Cardenas A.A. et al. Limiting the impact of stealthy attacks on industrial control systems. *Proc. ACM Conf. on Computer and Communications Security (CCS)*, 2016, pp. 1092–1105.

[5] Liu Y., Reiter M.K., Ning P. False data injection attacks against state estimation in electric power grids. *ACM Trans. Inf. Syst. Secur.*, 2011, vol. 14, no. 1, pp. 13:1–13:33.

[6] Chen J., Patton R. Fault diagnosis in dynamic systems: theory and application. *Prentice Hall*, 1999. 345 p.

[7] Gertler J. Fault detection and diagnosis in engineering systems. *CRC Press*, 1998. 493 p.

[8] Ding S.X. Model-based fault diagnosis techniques: design schemes, algorithms and tools. *Springer*, 2008. 303 p.

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

[10] Zhang Y., Jiang J. Bibliographical review on reconfigurable fault-tolerant control systems. *Annual Reviews in Control*, 2008, vol. 32, no. 2, pp. 229–252.

[11] Новіков О., Ільїн М., Стьопочкіна І., Овчарук М. Визначення параметрів непомітних кібератак на системи керування об’єктів критичної інфраструктури // KPI Science News - No. 1, 2025, p. 69-75.

[12] L. Alekseichuk, O. Novikov, A. Rodionov, D. Yakobchuk Cyber Security Logical and Probabilistic Model of a Critical Infrastructure Facility in the Electric Energy Industry // Theoretical And Applied Cybersecurity - Vol.5 No. 1, 2023, p. 61-66.

[13] Novikov O., Shreider M., Stopochkina I., Ilin M. Cyber Attacks Simulation for Modern Energy Facilities // CEUR Workshop Proceedings. – 2023. – Vol. 3887. – C. 35–49.

[14] Bryson A.E., Ho Y.C. *Applied Optimal Control: Optimization, Estimation, and Control*. Taylor & Francis, 1975.

[15] Bertsekas D.P. *Nonlinear Programming*. Athena Scientific, 1999.

[16] Kirk D.E. *Optimal Control Theory: An Introduction*. Dover Publications, 2004.

[17] Sahinidis N.V. Optimization under uncertainty: state-of-the-art and opportunities. *Computers & Chemical Engineering*, 2004, vol. 28, no. 6–7, pp. 971–983.