Breaking the Flow and the Bank:
Stealthy Cyberattacks on Water Network Hydraulics

Abdallah Alalem Albustami[a], Ahmad F. Taha[a]

[a]*Vanderbilt University, Civil and Environmental Engineering Department, Nashville, 37235, TN, US*

Abstract

As water distribution networks (WDNs) become increasingly connected with digital infrastructures, they face greater exposure to cyberattacks that threaten their operational integrity. Stealthy False Data Injection Attacks (SFDIAs) are particularly concerning, as they manipulate sensor data to compromise system operations while avoiding detection. While existing studies have focused on either detection methods or specific attack formulations, the relationship between attack sophistication, system knowledge requirements, and achievable impact remains unexplored. This paper presents a systematic analysis of sensor attacks against WDNs, investigating different combinations of physical constraints, state monitoring requirements, and intrusion detection evasion conditions. We propose several attack formulations that range from tailored strategies satisfying both physical and detection constraints to simpler measurement manipulations. The proposed attacks are simple and local—requiring knowledge only of targeted sensors and their hydraulic connections—making them scalable and practical. Through case studies on Net1 and Net3 benchmark networks, we demonstrate how these attacks can persistently increase operational costs and alter water flows while remaining undetected by monitoring systems for extended periods. The analysis provides utilities with insights for vulnerability assessment and motivates the development of protection strategies that combine physical and statistical security mechanisms.

Keywords: Stealthy cyber–physical attacks, False data injection attacks, Water distribution systems, State-estimation.

1. Introduction

Water Distribution Networks (WDNs) are undergoing a significant transformation as cities evolve towards smarter infrastructure. This change from traditional, centralized systems to more distributed intelligent networks brought tremendous improvements in real-time decision making and resource optimization. It has, nevertheless exposed vulnerabilities in cybersecurity of water systems.

In recent years, the water sector has experienced several notable cyberattacks. In 2021, a cyberattack targeted a water treatment facility in Oldsmar, Florida, attempting to poison the water supply by altering chemical levels [14]. More recently, in April 2024, a hacking group targeted multiple water utilities in Texas, causing a water tank to overflow and disrupting the SCADA systems that control hydraulic operations [22]. Despite such incidents, many cyberattacks go unnoticed or are undisclosed due to concerns over reputation and customer trust [17].

A common cyberattack on WDNs occurs when adversaries gain access to the operational technology (OT) network, typically by infiltrating the information technology (IT) network via methods like phishing, ransomware, or exploiting security vulnerabilities. Once inside the OT network, attackers can manipulate critical components such as SCADA systems, programmable logic controllers (PLCs), and sensors, allowing them to alter operational parameters [9]. One particularly dangerous and common form of such attacks is a False Data Injection Attack (FDIA). In an FDIA, attackers manipulate sensor readings or actuator data to mislead state monitoring algorithms into accepting falsified values as legitimate. This is concerning in WDNs because these attacks can be *stealthy*, bypassing intrusion detection (ID) methods and potentially causing significant operational and physical damage before being noticed by operators [44]. While FDIAs have been extensively studied in fields like smart grids [7, 19] transportation networks [4, 5], and the control engineering literature [30, 48, 49], they have received less attention in WDNs, despite their potential to severely disrupt essential services and compromise public safety [26]. The unique hydraulic characteristics of these systems necessitate tailored approaches to FDIA formulation and detection, which remain understudied. This gap motivates our work on addressing the specific challenges posed by FDIA in WDNs. The following section reviews the current state of FDIA research in this domain.

1.1. Literature Review

The Battle of the Attack Detection Algorithms (BATADAL) marked a significant milestone in developing and testing effective detection methods for cyber-physical attacks on WDNs [39]. The competition featured seven algorithms for detecting cyber-physical attacks, with the only model-based approach demonstrating superior performance [18]. The attacks in BATADAL were simulated using epanetCPA, a MATLAB

Email addresses: `abdallah.b.alalem.albustami@vanderbilt.edu` (Abdallah Alalem Albustami), `ahmad.taha@vanderbilt.edu` (Ahmad F. Taha)

modeling toolbox that enables the simulation of hydraulic responses to cyber-physical attacks via EPANET [38]. As such, these attacks were not FDIAs nor were they designed with stealthiness in mind. Following BATADAL, some researchers have developed more advanced attack detection algorithms and investigated different strategies, often utilizing similar simulation environments [12, 16, 27, 34].

Despite the growing interest in WDN security, studies specifically addressing FDIAs and their stealthy variants remain limited. Urbina *et al.* [44] analyzed stealthy attacks in water treatment systems, finding that actuator attacks are harder to launch than sensor attacks, and that detectors using historical data (stateful) perform better than those examining individual data points (stateless). However, their work focused on pH sensors and pumps in water treatment, not addressing WDN hydraulics. Douglas *et al.* [15] extended the epanetCPA toolkit to simulate cyber-physical attacks by interrupting sensor readings to impact water levels in tanks, providing insights into hydraulic impacts but not specifically formulating FDIAs. In [2], a water distribution testbed (WADI) was developed using random false data injections into tank level sensors. Although their work enabled experimental assessment, the random nature of these attacks suggests they might be detectable through conventional methods such as simple ID algorithms and residual checks during state estimation (SE) processes.

More relevant to this work are the few studies that have directly addressed SFDIAs against WDN hydraulics. Ahmed *et al.* [1] presented a case study on a model-based detection approach for smart WDNs. They utilized a Kalman filter for SE and compared a CUmulative SUM (CUSUM) statistic with a Bad Data Detector (BDD) for ID. They evaluated three attack scenarios, including Bias Injection Attacks (simple FDIAs that add constant offsets to measurements), Zero-Alarm Attacks (similar to SFDIAs but focused only on detection bypass), and attacks on control inputs. Their analysis of detection difficulty—showing that deceiving CUSUM is more complex than bypassing BDD due to its accumulated sum over time—aligns with findings from other domains, as their approach did not specifically address WDN hydraulics or consider physical constraints that could make attacks implausible due to mass and energy balance violations.

Exploring the offensive side of SFDIAs, Moazeni and Khazaei [25] proposed a nonlinear programming framework for modeling SFDIAs on flow rate measurements and total demand, targeting storage tanks through a bi-level optimization approach. They later adapted this strategy to target pump flow rate measurements, aiming to exceed maximum pressure heads at multiple nodes [24]. While their work demonstrated the feasibility of attacks that satisfy hydraulic constraints while bypassing state estimation and bad data detection, it was limited to specific scenarios and detection methods. In this work, we develop a broader analysis encompassing multiple detection mechanisms and attack formulations, from random to coordinated manipulations, with varying operational targets. Raza and Moazeni [35] introduced a robust chance-constrained optimization strategy to identify vulnerable locations in Smart WDNs against SFDIAs, considering the probabilistic nature of water

demand at junctions.

While the reviewed literature demonstrates growing attention to FDIAs in WDNs, most studies focused on specific attack scenarios or detection methods in isolation. To comprehensively understand system vulnerabilities, it was essential to analyze attacks across the full spectrum—from tailored worst-case scenarios to simpler, more common threats. This analysis required examining two critical components in cyber-physical systems: State Estimation (SE) and Intrusion Detection (ID). SE infers the system's overall state from sensor measurements, while ID analyzes discrepancies between measurements and state estimates to identify anomalies. A common thread across cyber-physical security literature is that effective FDIAs must evade both SE and ID mechanisms to remain stealthy. In power systems [21, 46] and water networks alike [24, 25], attackers additionally need to ensure that manipulated measurements satisfy domain-specific physical constraints—power flow equations in electrical grids and hydraulic relationships in water systems. The control engineering literature offers more generalized attack frameworks [30, 48] that abstract these principles, but our approach specifically addresses the hydraulic constraints of WDNs while maintaining the essential stealth properties identified across domains. This enables practical vulnerability assessment in water infrastructure where the combination of physical laws and detection algorithms creates a multi-layered security challenge. Although SE had been extensively researched and implemented in power grids and other cyber-physical domains, its application in pressurized WDNs remains an active area of research despite the widespread adoption of SCADA systems. This is mainly due to sparse sensor placement and a lack of accurate, calibrated models. To that end, we analyze different classes of sensor attacks, considering varying levels of system knowledge and security settings.

1.2. Paper Contributions

This paper investigates different classes of sensor attacks targeting WDNs. We use a Weighted Least Squares (WLS) method for state estimation (SE) and two intrusion detection (ID) methods: Cumulative Sum (CUSUM) detector and the Chi-squared detector. We formulate several SFDIA strategies with varying objectives assumptions regarding system knowledge. The main contributions of this work are summarized as follows:

- Formulation and analysis of sensor attacks against WDNs by analyzing interactions between physical constraints (mass/energy balance), state estimation convergence requirements, and intrusion detection evasion. From this analysis emerges four distinct attack strategies corresponding to different attacker capabilities and system vulnerabilities, providing a foundation for understanding the relationship between an attacker's system knowledge and their ability to manipulate measurements while avoiding detection.
- Quantitative evaluation of how different detection approaches—dynamic CUSUM monitoring, static chi-squared tests, and physical validation checks—perform against these attacks, along with analysis of their impacts on hydraulic operations through pump scheduling and

water flow management. The results highlight how local hydraulic relationships constrain attack capabilities but also enable targeted manipulations that can significantly increase operational costs and alter water flows while maintaining both physical and detection stealthiness.

- Development of a systematic methodology for strategic measurement selection and attack execution, including: *(i)* identification of vulnerable network configurations requiring minimal sensor manipulations, *(ii)* analysis of how network topology (radial vs. looped) affects attack complexity, and *(iii)* creation of a comprehensive algorithm that guides practitioners through the entire attack implementation process from target selection to impact assessment. This enables utilities to systematically evaluate security vulnerabilities without requiring specialized expertise in optimization theory.

The broader impact of this work is providing utilities with a comprehensive understanding of potential vulnerabilities across different security configurations, supporting proactive defense planning against the growing spectrum of cyber threats.

1.3. Paper Organization

The remainder of this paper is structured as follows: Section 2 presents the preliminaries, including the WDN hydraulic model, SE approaches, and ID methods. Section 3 develops the full-stealth attack formulation, and Section 4 presents additional attack strategies. Section 5 evaluates these attacks through case studies. Section 6 provides a systematic approach to attack design and implementation, and Section 7 discusses limitations and future research directions. Finally, Section 8 concludes the paper.

2. Preliminaries

This section introduces the hydraulic model for the studied water distribution network, state estimation, and the intrusion detection methods utilized in this study.

2.1. Hydraulic Modeling of Water Distribution Networks

In this section, we describe the hydraulic modeling of Water Distribution Networks (WDNs), which consists of reservoirs, tanks, junctions, pipes, pumps, and valves. The hydraulic behavior of these components is governed by the principles of conservation of mass and energy. Each of these components is described below along with their governing equations.

2.1.1. Reservoirs

Reservoirs in the network are considered as infinite water sources with a constant hydraulic head, representing a fixed elevation [47]. The head at Reservoir i at time step k is given by: $h_i^R(k + 1) = h_i^R(k)$, where $h_i^R(k)$ is the head at Reservoir i at time step k and remains constant over time. This assumption is valid as reservoirs are often located at a high elevation and supply water under gravity.

2.1.2. Tanks

Tanks are modeled with constant cross-sectional areas, and their head changes dynamically depending on the balance of inflows and outflows. The head at Tank i at time step $k + 1$ is

given by:

$$h_i^{TK}(k + 1) = h_i^{TK}(k) + \frac{\Delta t}{A_i^{TK}} \left(\sum_{j \in L_{\text{in}}} Q_{ij}^{\text{in}}(k) - \sum_{k \in L_{\text{out}}} Q_{ik}^{\text{out}}(k) \right), \quad (1)$$

where $h_i^{TK}(k)$ is the head at Tank i at time step k, A_i^{TK} is the cross-sectional area of the tank, $Q_{ij}^{\text{in}}(k)$ is the flow into the tank through connected links j, and $Q_{ik}^{\text{out}}(k)$ is the outflow from the tank through links k. The head variation is directly proportional to the net water flow into or out of the tank.

2.1.3. Junctions

Junctions in a WDN represent intersections of water flow, where the conservation of mass ensures that the total inflows equal the total outflows plus any local demand at the junction. The mass balance at Junction i is described by:

$$\sum_{j \in L_{\text{in}}} Q_{ij}^{\text{in}}(k) = \sum_{k \in L_{\text{out}}} Q_{ik}^{\text{out}}(k) + Q_i^D(k), \quad (2)$$

where $Q_i^D(k)$ represents the demand at Junction i at time step k, $Q_{ij}^{\text{in}}(k)$ is the inflow from links connected to i, and $Q_{ik}^{\text{out}}(k)$ is the outflow to links connected to i. The equation ensures that the water entering a junction is balanced by the water leaving and the demand at that junction.

2.1.4. Pipes

Water flowing through pipes experiences head losses due to friction and minor losses (e.g., bends). These head losses are computed using the Hazen-Williams formula, which models the relationship between flow and head loss as:

$$\Delta h_i(k) = h_j(k) - h_k(k) = r_i \, Q_i(k) \, |Q_i(k)|^{\mu - 1}, \quad (3)$$

where $\Delta h_i(k)$ is the head loss across Pipe i, $h_j(k)$ and $h_k(k)$ are the heads at Nodes j and k, respectively, $Q_i(k)$ is the flow through the pipe, and r_i is the pipe resistance coefficient, given by $r_i = \frac{4.727 \times L_i}{C_i^{1.852} \times D_i^{4.8704}}$, where L_i is the length of the pipe, C_i is the roughness coefficient (typically between 100-140), and D_i is the pipe diameter. The flow exponent, μ, is set to 1.852, consistent with the Hazen-Williams equation [20]. It is worth noting that while specific hydraulic formulas are presented here, the methods developed in this paper remain applicable to any hydraulic modeling approach as they rely only on fundamental mass and energy conservation principles.

2.1.5. Pumps

Pumps add energy to the water flow, increasing the head between the upstream and downstream nodes [20]. The head gain provided by Pump i, connecting Nodes j and k, is modeled by:

$$\Delta h_i^M(k) = h_j(k) - h_k(k) = -s_i^2(k) \left(h_i^0 - \alpha_i \left(s_i^{-1}(k) Q_i^M(k) \right)^\nu \right), \quad (4)$$

where $s_i(k)$ is the relative speed of the pump, h_i^0 is the shutoff head (the maximum head when there is no flow), $Q_i^M(k)$ is the flow through the pump, α_i and ν are pump-specific coefficients, and $h_j(k)$ and $h_k(k)$ are the heads at the upstream and downstream nodes, respectively. The speed $s_i(k)$ ranges between 0 and the maximum speed s_i^{max}, determining the pump's operating speed.

2.1.6. Valves

Valves are used to control flow in the network, and in this model, they are considered as on-off components. A valve can either be fully open, allowing water to flow as in a regular pipe, or fully closed, decoupling the two connected nodes. For an open valve, the head loss across Valve i connecting Nodes j and k is given by:

$$\Delta h_i^V(k) = h_j(k) - h_k(k) = m_i Q_i^V(k)|Q_i^V(k)|, \tag{5}$$

where $Q_i^V(k)$ is the flow through the valve, and m_i is the minor loss coefficient associated with the valve. When the valve is closed, no flow occurs, and the two nodes are effectively decoupled.

We employ a piecewise linearization approximation for the nonlinear hydraulic components following the approach in [23]. For pipes and valves, each head loss curve is segmented into linear pieces determined by connecting points calculated offline. For a pipe connecting nodes i and j, the linearized head loss is represented through:

$$h_{j_k} - h_{i_k} - \sum_{n=1}^{N_{PW}} m_n \zeta_{n_k} - \sum_{n=1}^{N_{PW}} b_n \omega_{n_k} = 0, \tag{6a}$$

$$q_{i_k} - \sum_{n=1}^{N_{PW}} \zeta_{n_k} = 0, \tag{6b}$$

$$\sum_{n=1}^{N_{PW}} \omega_{n_k} = 1, \tag{6c}$$

$$-\zeta_{n_k} + q_{n,\min} \omega_{n_k} \leq 0, \tag{6d}$$

$$\zeta_{n_k} - q_{n,\max} \omega_{n_k} \leq 0, \tag{6e}$$

where m_n and b_n represent the slope and intercept of segment n, ζ_{n_k} is the flow through segment n, and ω_{n_k} is a binary variable selecting segment n. Equation (6a) defines the piecewise linear head loss, (6b) ensures flow conservation across segments, (6c) enforces single segment selection, (6d) and (6e) constrain flows within segment bounds.

For pumps, following [28], we approximate the characteristic curves using quadratic functions:

$$\Delta h_{i_k}^M = \beta_1 (q_{i_k}^M)^2 + \beta_2 q_{i_k}^M + \beta_3 (s_{i_k}^M)^2 + \beta_4 \tag{7}$$

where coefficients β_1–β_4 are determined by minimizing approximation error while ensuring convexity through $\beta_1, \beta_3 \geq 0$. This approximation maintains the relationship between pump speed, discharge, and head gain while enabling computationally tractable optimization formulations.

2.2. Hydraulics State Estimation in WDNs

The literature on SE for hydraulics of WDNs covers a wide range of methods. Static methods, which process measurements independently at each time step, aim to minimize residuals through iterative techniques such as sum of absolute or squared errors [6, 8]. Other approaches focus on uncertainty bounds through confidence limit analysis and interval hydraulic SE [10, 45]. In contrast, dynamic methods like the extended Kalman filter [11, 36] account for system evolution over time

and have recently gained traction. SE provides system operators with several advantages: determining initial system state, providing real-time snapshots and detecting intrusions or anomalies [31, 42]. However, its adoption remains limited due to requirements for well-calibrated models and sufficient sensor placement for network observability. This work implements a weighted least squares (WLS) approach for state estimation, though the analysis principles remain valid for other SE methods that can provide comparable state estimates.

2.2.1. Weighted Least Squares State Estimation

Let $x_k \in \mathbb{R}^n$ represent the state vector at time step k, where n is the number of state variables in the system, including flows at pipes and heads at junctions. The measurement vector is denoted by $y_k \in \mathbb{R}^m$, where m represents the number of available sensor measurements, such as flow rates and pressures at selected nodes and pipes in the network. The relationship between the measurements and the system state is described by the equation $y_k = h(x_k) + v_k$, where $h(x) = Hx$ is the linear measurement function with $H \in \mathbb{R}^{m \times n}$ being the measurement matrix, and $v_k \in \mathbb{R}^m$ represents the measurement noise, assumed to be Gaussian with zero mean and known covariance. The objective of WLS is to minimize the weighted sum of squared residuals between the measured values y_k and the predicted measurements $h(\hat{x})$ [6]. The WLS problem is formulated as:

$$\underset{\hat{x}_k}{\text{minimize}} (y_k - H\hat{x}_k)^\top W(y_k - H\hat{x}_k), \tag{8}$$

which, given a linear measurement function, has the following analytical solution:

$$\hat{x}_k = (H^\top W H)^{-1} H^\top W y_k, \tag{9}$$

where $W \in \mathbb{R}^{m \times m}$ is a diagonal weight matrix, with entries representing the inverse of the measurement variances.

While in our implementation we do not solve a formal sensor-placement optimization problem, we distribute flow and pressure sensors so that the system is observable via a WLS-based approach, ensuring the operator can reconstruct the relevant states. For measurement noise, we assume each sensor has a known variance and incorporate standard deviations directly into the weight matrix with each element representing the inverse variance of the corresponding measurement. The specific configuration of the WLS setup in this work can be found in Appendix A.

2.3. Intrusion Detection

An Intrusion Detection (ID) algorithm identifies anomalies through examining the measurement residuals, defined as the differences between sensor measurements and the state estimates, $r_k \in \mathbb{R}^m$:

$$r_k = y_k - H\hat{x}_k, \tag{10}$$

Two distinct detection mechanisms are used in this study: the Cumulative Sum (CUSUM) detector and the Chi-squared (χ^2) detector.

2.3.1. CUSUM Detection

The Cumulative Sum (CUSUM) detector is a dynamic detection method that tracks changes in the cumulative sum of the residuals over time [29]. The detector works by accumulating deviations in the residuals from expected values, making it well-suited for identifying persistent, stealthy anomalies. While CUSUM has proven effective in various domains, its application in WDNs remains relatively limited, with a few studies applying it for water quality monitoring [44] and hydraulic anomaly detection [1].

The CUSUM detection process operates by comparing a cumulative statistic c_k to a bias term b and a predefined threshold τ. The CUSUM procedure is defined as follows:

$$c_1 = 0, \quad c_k = \begin{cases} \max(0, c_{k-1} + z_k - b), & \text{if } c_{k-1} \leq \tau, \\ 0 \text{ and } \tilde{k} = k - 1, & \text{if } c_{k-1} > \tau. \end{cases} \quad (11)$$

where z_k represents the deviation from normal operation (distance measure), and $b \in \mathbb{R}_{>0}$ is the bias term that adjusts the detector's sensitivity. An alarm is triggered when c_k exceeds a predefined threshold $\tau \in \mathbb{R}_{>0}$, signaling the detection of an anomaly. Once the alarm is triggered, c_k is reset to zero, and the process continues. The CUSUM can be implemented in two ways: scalar or vectorized. The scalar approach uses $z_k = r_k^\top \Sigma^{-1} r_k$ to detect anomalies through collective evaluation of residuals, while the vectorized approach uses $z_k = |r_k|$ to monitor residuals independently, enabling detection of localized anomalies. Although [28] suggests a theoretical framework for optimal CUSUM parameter tuning, this work empirically adjusts threshold and bias parameters based on historical data to acheive the desired false alarm rate.

2.3.2. Chi-squared (χ^2) Detector

The Chi-squared detector is a static detection mechanism designed to identify sudden anomalies through checking whether the residual vector at time step k, denoted by r_k, falls within the expected distribution [13]. The detection process uses a chi-squared test statistic, z_k, defined as:

$$\text{If } z_k = r_k^\top \Sigma^{-1} r_k > \alpha, \quad \tilde{k} = k. \quad (12)$$

where Σ^{-1} is the inverse of the residual covariance matrix. An alarm is triggered when the test statistic exceeds a predefined threshold α. This threshold is computed using the inverse regularized lower-incomplete gamma function $P^{-1}(\cdot)$, ensuring that the detection algorithm maintains a desired false alarm rate. Specifically, α is calculated as $\alpha = 2P^{-1}\left(\frac{n_y}{2}, 1 - \frac{1}{\gamma}\right)$, where n_y is the number of independent measurements, and γ is the desired mean time between false alarms.

Both detection methods offer complementary capabilities for identifying cyber-physical attacks. The CUSUM detector is better suited for detecting subtle, long-term deviations, and the Chi-squared detector identifies sudden anomalies that fall outside the expected range of residual values.

Next, we present a set of tailored sensor attacks that manipulate flow, pressure, and demand measurements. The attacks are formulated with varying levels of constraint satisfaction and system knowledge requirements. Fig. 1 illustrates the WDN architecture and its security components, showing how SFDIAs
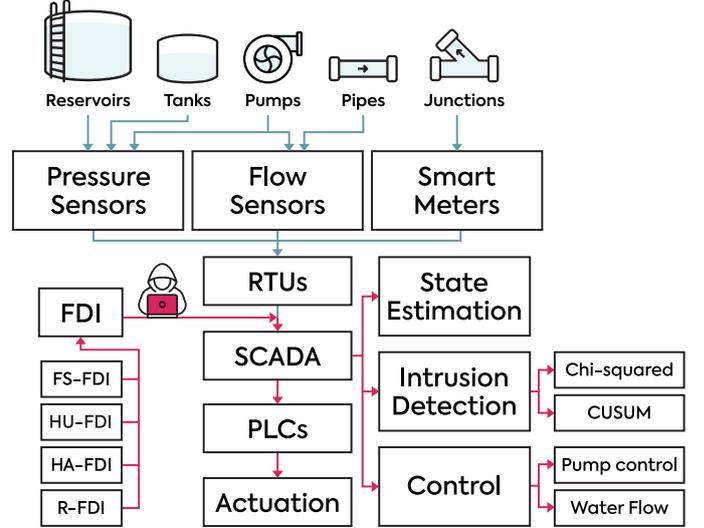

Figure 1: Water distribution network architecture illustrating system components, monitoring systems, and security mechanisms. Red arrows indicate attack vectors, demonstrating how different SFDIA variants (FS-FDI: Full-Stealth FDI, HU-FDI: Hydraulics-Unaware FDI, HA-FDI: Hydraulics-Aware FDI, R-FDI: Random FDI) interact with the system's security mechanisms.

target various system layers. Tab. 1 defines the key variables and notations used throughout the analysis. Tab. 2 summarizes the attack strategies, comparing their characteristics and required knowledge.

3. Full-Stealth Attack Design

This section presents an optimization-based attack formulation that simultaneously satisfies physical constraints and bypasses detection mechanisms. The Full-Stealth FDI (FS-FDI) strategy is designed to: (i) evade intrusion detection mechanisms, (ii) maintain state estimation convergence, and (iii) satisfy hydraulic constraints governed by mass and energy balance principles. We formally define the FS-FDI strategy through the following nonlinear optimization problem:

$$\underset{a_k^h, a_k^f, a_k^d}{\text{maximize}} \quad \sum_{i=1}^{n_h} |a_{i,k}^h| + \sum_{j=1}^{n_f} |a_{j,k}^f| + \sum_{l=1}^{n_d} |a_{l,k}^d| \quad (13a)$$

$$\text{subject to:} \quad \hat{x}_k^a = \mathcal{F}_a(y_k^a, W_a), \quad (13b)$$

$$y_k^{a,h} = y_k^h + a_k^h, \quad (13c)$$

$$y_k^{a,f} = y_k^f + a_k^f, \quad (13d)$$

$$d_k^a = d_k + a_k^d, \quad (13e)$$

$$\mathcal{M}(\hat{x}_k^a, d_k^a) = 0, \quad (13f)$$

$$\mathcal{E}(\hat{x}_k^a) = 0, \quad (13g)$$

$$r_k^a = [y_k^{a,h}; y_k^{a,f}] - h_a(\hat{x}_k^a), \quad (13h)$$

$$\mathcal{D}(r_k^a) \leq \beta, \quad (13i)$$

$$\|\hat{x}_k^a - x_{\text{ref}}^a\| \leq \epsilon, \quad (13j)$$

$$\|a_k^h\|_\infty \leq \alpha_h \|y_k^h\|_\infty, \quad (13k)$$

$$\|a_k^f\|_\infty \leq \alpha_f \|y_k^f\|_\infty, \quad (13l)$$

$$\|a_k^d\|_\infty \leq \alpha_d \|d_k\|_\infty \quad (13m)$$

Table 1: Notation for Attack Formulation

Symbol	Description
$\boldsymbol{x}_k \in \mathbb{R}^n$	True state vector (flows and heads)
$\hat{\boldsymbol{x}}_k \in \mathbb{R}^n$	Estimated state vector
$\boldsymbol{y}_k \in \mathbb{R}^m$	Original measurement vector
$\boldsymbol{y}_k^{a,h} \in \mathbb{R}^{n_h}$	Attacked head measurements
$\boldsymbol{y}_k^{a,f} \in \mathbb{R}^{n_f}$	Attacked flow measurements
$\boldsymbol{d}_k^a \in \mathbb{R}^{n_d}$	Attacked demand measurements
$\boldsymbol{a}_k^h \in \mathbb{R}^{n_h}$	Head measurement attack vector
$\boldsymbol{a}_k^f \in \mathbb{R}^{n_f}$	Flow measurement attack vector
$\boldsymbol{a}_k^d \in \mathbb{R}^{n_d}$	Demand attack vector
$\mathcal{M}, \mathcal{E}$	Mass and energy balance functions
$\mathcal{D}$	Detection function
$\boldsymbol{H}_a \in \mathbb{R}^{m \times n}$	Measurement matrix for attacked subsystem
$\boldsymbol{W}_a \in \mathbb{R}^{m \times m}$	Weight matrix for attacked measurements
$\boldsymbol{a}_k^{\text{drift}} \in \mathbb{R}^m$	Random walk drift component
$\boldsymbol{a}_k^{\text{noise}} \in \mathbb{R}^m$	High-frequency noise component
$\boldsymbol{a}_k^{\text{spike}} \in \mathbb{R}^m$	Occasional spike component

The objective function in (13a) minimizes the sum of absolute attack values, where $\boldsymbol{a}_k^h \in \mathbb{R}^{n_h}$, $\boldsymbol{a}_k^f \in \mathbb{R}^{n_f}$, and $\boldsymbol{a}_k^d \in \mathbb{R}^{n_d}$ represent attack vectors for head, flow, and demand measurements respectively, with n_h, n_f, and n_d being the number of targeted sensors. This formulation promotes sparse attacks that target specific measurements, aiming to make the attack harder to detect through manual inspection or additional security measures beyond standard intrusion detection.

The state estimation constraint (13b) represents the attacker's implementation of SE over the attacked measurements and their hydraulically connected components. In the case of considering WLS, the function $\mathcal{F}_a$ simply represents $\hat{\boldsymbol{x}}_k^a = (\boldsymbol{H}^\top \boldsymbol{W} \boldsymbol{H})^{-1} \boldsymbol{H}^\top \boldsymbol{W} \boldsymbol{y}_k^a$. The convergence constraint (13j) ensures that the attacked states remain within reasonable bounds by limiting their deviation from reference states $\boldsymbol{x}\text{ref}^a$ observed during normal operation. Together, these constraints guarantee that operator-side state estimation will converge under attack while requiring knowledge of only the relevant network portion. We would also like to note that in our implementation, if an operator *does* employ a state estimation (SE) scheme that yields reliable estimates (so that measured values can be compared against these estimates for inrtusion detection), it does not fundamentally matter how that SE process is implemented—so long as an attacker has access to or can infer the estimates for the specific measurements they wish to manipulate. This means that regardless of whether the operator uses WLS, an Extended Kalman Filter (EKF), or any other SE approach, an attacker who can learn or approximate the relevant parameters (or simply obtain the state estimates for the sensors under attack) can design manipulations that remain undetected. To implement these constraints, an attacker needs the measurement equations and weights used in the operator's SE algorithm, along with typical state bounds. Such information can be inferred from SCADA system configurations and historical operating data during reconnaissance of the OT network.

The mass balance constraint (13f) enforces flow conservation at nodes using state vector components $\hat{\boldsymbol{x}}_k^a = [\boldsymbol{h}_k^a; \boldsymbol{q}_k^a]$ and altered demands $\boldsymbol{d}_k^a$ following the same flow conservation principles in (2). For each node $i \in \mathcal{N}$, function $\mathcal{M}$ ensures:

$$\sum_{j \in L_{\text{in}}} q_{ij_k}^a = \sum_{k \in L_{\text{out}}} q_{ik_k}^a + d_{i_k}^a \tag{14}$$

where $q_{ij_k}^a$ and $q_{ik_k}^a$ represent flows into and out of node i, and $d_{i_k}^a$ is the altered demand. In practice, this constraint need only be enforced within the local subnetwork affected by the attack, defined by the hydraulically connected components around targeted sensors. An attacker can identify these components through analysis of SCADA data to trace flow patterns and determine pipe connectivity.

The energy balance constraint (13g) maintains consistency in the hydraulic grade line across attacked measurement paths, following the relationships established in (3) for pipes and (4) for pumps. For connected nodes i and j, function $\mathcal{E}$ enforces:

$$h_{j_k}^a - h_{i_k}^a = \Delta h_{ij_k}(\boldsymbol{q}_k^a, \boldsymbol{s}_k^a) \tag{15}$$

where Δh_{ij_k} captures the appropriate head loss/gain based on the component type, using the piecewise linearization approach in (6). This constraint needs only be satisfied along paths containing attacked measurements, requiring the attacker to know pipe parameters and pump characteristics of the local subsystem.

Measurement residuals are computed in (13h), where $\boldsymbol{h}_a(\cdot)$ is a measurement function relating states to measurements. Constraint (13i) ensures the attack bypasses the intrusion detection mechanism. For a CUSUM detector with vectorized distance measure, $\mathcal{D}(\boldsymbol{r}_k^a) = \boldsymbol{c}_{k-1} + |\boldsymbol{r}_k^a| - \boldsymbol{b}_a \leq \boldsymbol{\tau}_a$, where the attacker only needs to know detector parameters (b_i and τ_i) corresponding to targeted sensors. For a chi-squared detector or CUSUM with scalar distance measure, $\mathcal{D}(\boldsymbol{r}_k^a) = \boldsymbol{r}_k^{a\top} \Sigma_a^{-1} \boldsymbol{r}_k^a \leq \alpha$, the residuals are aggregated through a weighted sum requiring knowledge of the full covariance matrix Σ_a and threshold α.

Attack magnitude constraints (13k)–(13m) limit attacks proportionally to original measurement magnitudes through coefficients α_h, α_f, and α_d. The infinity norm ensures no single attack component exceeds these bounds.

The FS-FDI optimization problem, with piecewise linearized hydraulic constraints (6), takes the form of a mixed-integer linear program (MILP). While such problems are NP-hard in general, the localized nature of the attack—considers targeted components and their immediate hydraulic connections only—keeps the problem size manageable and computationally tractable regardless of the overall network size. The relatively slow dynamics of WDN operations provide sufficient time for solving the optimization at each time step using standard MILP solvers like Gurobi. Even with hydraulic time steps as small as one minute, which is more frequent than typical operational requirements, the optimization remains computationally feasible.

The implementation of the FS-FDI strategy follows Algorithm 1, which iteratively solves the optimization problem in (13) while ensuring both algorithmic stealthiness and physical feasibility at each time step. The algorithm requires comprehensive system knowledge including target measurements, cur-

Table 2: Comparison of FDI attack strategies highlighting implementation requirements and analytical objectives.

Attack Strategy	Constraint Satisfaction	Required Knowledge	Technical Description	Analysis Objective
Full-Stealth FDI (FS-FDI) (13)	Mass/energy balance, SE convergence, ID bypass	ID parameters, local topology and hydraulic parameters, SE configuration	Multi-constraint optimization with coupled measurement modifications	Worst-case stealthy attack analysis
Hydraulics-Unaware FDI (HU-FDI) (16), (17), (18)	SE convergence, ID bypass	ID parameters, SE configuration	Residual-based optimization with uncoupled measurements	Physical validation effectiveness
Hydraulics-Aware FDI (HA-FDI) (19)	Mass/energy balance	Local network topology and hydraulic parameters	Hydraulic constraint satisfaction problem	Detection mechanism assessment
Random FDI (R-FDI) (20)	None	Sensor operational bounds	Bounded measurement modifications	Base vulnerability metrics

rent state estimates, detector parameters, local network topology, and hydraulic parameters. At each iteration, the algorithm computes current residuals, solves for optimal attack vectors, and verifies that both detector statistics and physical constraints remain satisfied after measurement modification.

Algorithm 1 Full-Stealth FDI Implementation

Require: measurements $\boldsymbol{y}_k$, state estimates $\hat{\boldsymbol{x}}_k$, network topology and hydraulic parameters, detector parameters (τ, b, α), SE configuration and typical bounds
1: **for** each time step k **do**
2:     Compute current residuals: $\boldsymbol{r}_k = \boldsymbol{y}_k - \boldsymbol{h}(\hat{\boldsymbol{x}}_k)$
3:     Initialize attack vector variables $\boldsymbol{a}_k^h, \boldsymbol{a}_k^f, \boldsymbol{a}_k^d$
4:     Solve optimization problem (13)
5:     Update measurements:
6:         $\boldsymbol{y}_k^{a,h} = \boldsymbol{y}_k^h + \boldsymbol{a}_k^h$
7:         $\boldsymbol{y}_k^{a,f} = \boldsymbol{y}_k^f + \boldsymbol{a}_k^f$
8:         $\boldsymbol{d}_k^a = \boldsymbol{d}_k + \boldsymbol{a}_k^d$
9:     Verify detector statistics remain within thresholds
10:    Verify physical constraints are satisfied
11: **end for**
12: **return** modified measurements $[\boldsymbol{y}_k^{a,h}; \boldsymbol{y}_k^{a,f}]$ and demands $\boldsymbol{d}_k^a$

The FS-FDI strategy provides a general framework for analyzing worst-case stealthy sensor attacks against WDN hydraulics. The formulation ensures attack stealthiness through: *(i)* intrusion detection bypass, *(ii)* state estimation convergence, and *(iii)* physical constraint satisfaction. While the strategy requires significant system knowledge (see Tab. 2), it aligns with the literature on worst-case SFDIA formulation for cyber-physical systems. The proposed formulation serves several purposes. First, it provides a benchmark for evaluating system resilience against stealthy and tailored worst case sensor attacks, which are becoming increasingly common in critical infrastructure. Second, it enables operators to systematically identify vulnerabilities in their monitoring systems and develop targeted protection strategies. Third, despite its apparent complexity, the framework's reliance on standard hydraulic principles and

common operational data makes it practically implementable for both security testing and attack simulation. The larger time scales associated with WDN hydraulics provide flexibility for implementing these optimization-based attacks within feasible operational intervals.

While FS-FDI provides a complete attack formulation satisfying all security constraints, we next examine strategies that relax different combinations of these constraints to analyze how different constraints affect attack capabilities and impact.

4. Constraint-Relaxed Attack Strategies

This section presents three attack formulations with varying degrees of sophistication, examining how different constraint relaxations affect attack capabilities and required system knowledge.

4.1. Hydraulics-Unaware FDI (HU-FDI)

A Hydraulics-Unaware FDI attack is designed to maintain algorithmic stealthiness through ID bypass and SE convergence while neglecting the physical consistency requirements. The strategy enables analysis of how hydraulic constraints influence attack feasibility and helps quantify the trade-off between attack complexity and required system knowledge. Two main approaches are typically used in designing HU-FDI attacks, an optimization-based approach and a closed-form solution [43].

For the optimization approach, the HU-FDI strategy modifies the formulation in (13) by removing the hydraulic constraints (13f) and (13g), yielding a computationally simpler formulation that requires only detection parameters and targeted measurement information.

We also derive closed-form solutions for attack vectors that maintain detector statistics at their respective thresholds for both detection mechanisms presented in Section 2.3. These solutions take different forms depending on how the detector evaluates the deviation between measurements and estimates—either by examining each residual independently or by considering their collective weighted sum. For the vectorized CUSUM implementation where $z_k = |\boldsymbol{r}_k|$, the attack vector com-

ponents are designed independently [32, 43]:

$$a_{k,i} = \begin{cases} \pm(\tau_i + b_i - c_{k-1,i}) - r_{k,i}, & \text{if } k = k^* \\ b_i - r_{k,i}, & \text{if } k > k^* \end{cases} \tag{16}$$

where $i \in 1, \ldots, m$ indexes the targeted measurements. Here, the attacker only needs knowledge of individual residual parameters and measurements. For scalar distance measures, the closed-form solutions require more comprehensive system knowledge. The CUSUM detector with a scalar distance measure, $z = r_k^\top \Sigma^{-1} r_k$, yields [28]:

$$\boldsymbol{a}_k = \begin{cases} \Sigma^{\frac{1}{2}} \Gamma \left(\sqrt{\frac{\tau+b-c_{k-1}}{n}}, \ldots, \sqrt{\frac{\tau+b-c_{k-1}}{n}} \right)^\top - r_k, & \text{if } k = k^* \\ \Sigma^{\frac{1}{2}} \Gamma \left(\sqrt{\frac{b}{n}}, \ldots, \sqrt{\frac{b}{n}} \right)^\top - r_k, & \text{if } k > k^* \end{cases} \tag{17}$$

where Γ represents the selection matrix for targeted measurements. Similarly, for the chi-squared detector [28]:

$$\boldsymbol{a}_k = \Sigma^{\frac{1}{2}} \Gamma \left(\sqrt{\frac{\alpha}{n}}, \ldots, \sqrt{\frac{\alpha}{n}} \right)^\top - r_k, \tag{18}$$

For an ID mechanism that employs a scalar distance measure, an attacker implementing the closed-form solution would need comprehensive knowledge of the entire measurement vector, state estimates, residual covariance matrix, and detector parameters. This requirement stems from the scalar nature of the detection statistic, where residuals are collectively evaluated. In contrast, an optimization-based approach provides a more practical alternative, requiring only knowledge of the targeted measurements and their corresponding detector parameters. This makes the optimization-based approach more suitable for real-world implementation against scalar detection schemes.

While HU-FDI attacks can successfully evade detection, the manipulated measurements and resulting state estimates could violate physical laws, potentially alerting operators through obvious deviations from expected hydraulic behavior. This highlights the importance of incorporating physical constraints in attack design, as demonstrated by the FS-FDI strategy.

4.2. Hydraulics-Aware FDI (HA-FDI)

A Hydraulics-Aware FDI strategy represents attacks on systems where operators rely primarily on physical validation rather than intrusion detection mechanisms. This scenario is particularly relevant for WDNs where SE and ID systems may not be implemented due to cost or complexity constraints [41].

The HA-FDI strategy is a simplified version of the FS-FDI approach. It modifies the FS-FDI formulation in (13) by removing the ID and SE-related constraints in (13h)–(13j). By doing so, it focuses on satisfying physical constraints to ensure that the manipulated measurements remain hydraulically plausible. This approach allows the attacker to construct attack vectors that evade physical validation checks while requiring only knowledge of the local network topology and hydraulic parameters. The HA-FDI optimization is expressed as follows:

$$\underset{\boldsymbol{a}_k^h, \boldsymbol{a}_k^f, \boldsymbol{a}_k^d}{\text{maximize}} \quad \sum_{i=1}^{n_h} |a_{i,k}^h| + \sum_{j=1}^{n_f} |a_{j,k}^f| + \sum_{l=1}^{n_d} |a_{l,k}^d| \tag{19a}$$

subject to:
$$\boldsymbol{y}_k^{a,h} = \boldsymbol{y}_k^h + \boldsymbol{a}_k^h, \tag{19b}$$
$$\boldsymbol{y}_k^{a,f} = \boldsymbol{y}_k^f + \boldsymbol{a}_k^f, \tag{19c}$$
$$\boldsymbol{d}_k^a = \boldsymbol{d}_k + \boldsymbol{a}_k^d, \tag{19d}$$
$$\mathcal{M}(\hat{\boldsymbol{x}}_k^a, \boldsymbol{d}_k^a) = 0, \tag{19e}$$
$$\mathcal{E}(\hat{\boldsymbol{x}}_k^a) = 0, \tag{19f}$$
$$\|\boldsymbol{a}_k^h\|_\infty \le \alpha_h \|\boldsymbol{y}_k^h\|_\infty, \tag{19g}$$
$$\|\boldsymbol{a}_k^f\|_\infty \le \alpha_f \|\boldsymbol{y}_k^f\|_\infty, \tag{19h}$$
$$\|\boldsymbol{a}_k^d\|_\infty \le \alpha_d \|\boldsymbol{d}_k\|_\infty. \tag{19i}$$

This strategy provides a benchmark for assessing the limitations of detection systems that rely solely on physical validation checks. It highlights how attacks can bypass these checks by exploiting the hydraulic properties of the network. In practice, the HA-FDI strategy requires less computational effort compared to FS-FDI, as it avoids the need to account for ID or SE mechanisms. However, this also means that HA-FDI attacks may result in less algorithmically stealthy manipulations, potentially raising suspicion in systems equipped with detection systems.

4.3. Random FDI (R-FDI)

The Random FDI strategy represents attacks executed with minimal system knowledge and complexity, as a baseline for evaluating the effectiveness of both detection mechanisms and more advanced attack strategies. R-FDI implements structured randomization that reflects attack patterns that might emerge from automated scripts or basic manipulation tools.

We define the R-FDI attack through a multi-pattern random process that combines three attack components:

$$a_{i,k} = \begin{cases} a_{i,k}^{\text{drift}} + a_{i,k}^{\text{noise}} + a_{i,k}^{\text{spike}}, & \text{if } i \in \mathcal{T} \text{ and } k \ge k^* \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

The drift component introduces systematic bias through a random walk:

$$a_{i,k}^{\text{drift}} = a_{i,k-1}^{\text{drift}} + \delta_{i,k}, \quad \delta_{i,k} \sim \mathcal{N}(0, \sigma_d^2) \tag{21}$$

The noise component adds high-frequency perturbations:

$$a_{i,k}^{\text{noise}} = \nu_{i,k} \cdot \alpha_n \cdot y_{i,k}, \quad \nu_{i,k} \sim \mathcal{N}(0, \sigma_n^2) \tag{22}$$

The spike component introduces occasional large deviations:

$$a_{i,k}^{\text{spike}} = \begin{cases} s_{i,k} \cdot \alpha_s \cdot y_{i,k}, & \text{if } u_{i,k} \le p_s \\ 0, & \text{otherwise} \end{cases} \tag{23}$$

where $s_{i,k} \sim \mathcal{U}(-1, 1)$ generates random spike magnitudes, $u_{i,k} \sim \mathcal{U}(0, 1)$ determines spike occurrence with probability p_s, and α_s scales spike magnitude.

The final attack magnitude is constrained to maintain basic operational plausibility through:

$$a_{i,k}^{\text{final}} = \begin{cases} \alpha_{\max} |y_{i,k}|, & \text{if } a_{i,k} > \alpha_{\max} |y_{i,k}| \\ -\alpha_{\max} |y_{i,k}|, & \text{if } a_{i,k} < -\alpha_{\max} |y_{i,k}| \\ a_{i,k}, & \text{otherwise} \end{cases} \tag{24}$$

8

The implementation follows Algorithm 2. Fig. 2 presents a decision flowchart that guides attack strategy selection based on three key knowledge components: measurement access, understanding of security mechanisms (SE/ID), and network topology/parameters.

Algorithm 2 Random FDI Implementation

Require: Measurements $\boldsymbol{y}_k$, Target set $\mathcal{T}$, Parameters $\alpha_{\max}, \sigma_d, \sigma_n, \alpha_n, \alpha_s, p_s$
1: Initialize: $\boldsymbol{a}_0^{\text{drift}} = \boldsymbol{0}$
2: **for** each time step k **do**
3:     **for** each target $i \in \mathcal{T}$ **do**
4:         Generate $\delta_{i,k} \sim \mathcal{N}(0, \sigma_d^2)$
5:         Update drift: $a_{i,k}^{\text{drift}} = a_{i,k-1}^{\text{drift}} + \delta_{i,k}$
6:         Generate noise: $\nu_{i,k} \sim \mathcal{N}(0, \sigma_n^2)$
7:         Compute noise: $a_{i,k}^{\text{noise}} = \nu_{i,k} \cdot \alpha_n \cdot y_{i,k}$
8:         Generate $u_{i,k} \sim \mathcal{U}(0, 1)$
9:         **if** $u_{i,k} \leq p_s$ **then**
10:             Generate $s_{i,k} \sim \mathcal{U}(-1, 1)$
11:             Compute spike: $a_{i,k}^{\text{spike}} = s_{i,k} \cdot \alpha_s \cdot y_{i,k}$
12:         **else**
13:             Set $a_{i,k}^{\text{spike}} = 0$
14:         **end if**
15:         Combine components via (20)
16:         Apply bounds via (24)
17:         Update measurement: $y_{i,k}^a = y_{i,k} + a_{i,k}^{\text{final}}$
18:     **end for**
19: **end for**
20: **return** Modified measurements $\boldsymbol{y}_k^a$

4.4. Impact on Hydraulic Operations

We evaluate the proposed attacks' impact on two fundamental hydraulic operations: optimal pump scheduling and water flow. In the hydraulic model considered here, the overall system state $\boldsymbol{x} \in \mathbb{R}^n$ comprises the following physical quantities:

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{h}_j^\top & \boldsymbol{h}_t^\top & \boldsymbol{q}_p^\top & \boldsymbol{q}_{\text{pump}}^\top & \boldsymbol{s}_{\text{pump}}^\top \end{bmatrix}^\top \in \mathbb{R}^{n_j + n_t + n_p + 2n_m},$$

where $\boldsymbol{h}_j \in \mathbb{R}^{n_j}$ represents the hydraulic heads at junctions (in ft), $\boldsymbol{h}_t \in \mathbb{R}^{n_t}$ represents the hydraulic heads at tanks (in ft), $\boldsymbol{q}_p \in \mathbb{R}^{n_p}$ represents the flow rates in pipes (in GPM), $\boldsymbol{q}_{\text{pump}} \in \mathbb{R}^{n_m}$ represents the flow rates through pumps (in GPM), and $\boldsymbol{s}_{\text{pump}} \in \mathbb{R}^{n_m}$ represents the pump speeds (dimensionless, expressed as a fraction of maximum speed).

The pump scheduling optimization minimizes operational costs while maintaining hydraulic constraints through the following optimization:

$$\underset{\boldsymbol{s}, \boldsymbol{q}}{\text{minimize}} \quad \sum_{i=1}^{N_p} \varphi_{EL} \frac{\rho_w g}{\eta_i} q_i \Delta h_i^M \tag{25a}$$

$$\text{subject to:} \quad \text{Mass balance (2)} \tag{25b}$$

$$\text{Energy balance (3), (4)} \tag{25c}$$

$$\boldsymbol{s} \in [0, \boldsymbol{s}_{\max}], \quad \boldsymbol{q} \in [\boldsymbol{q}_{\min}, \boldsymbol{q}_{\max}] \tag{25d}$$

where φ_{EL} represents electricity price (\$/kWh), ρ_w is water density (kg/m^3), g is gravitational acceleration (m/s^2), η_i is
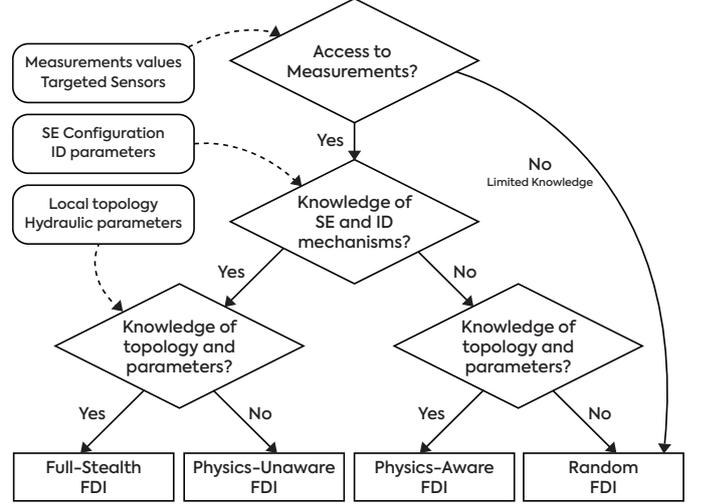

Figure 2: Decision-making flowchart for selecting FDI attack strategies based on available knowledge

pump efficiency (%), q_i is flow rate through pump i (m^3/s), and Δh_i^M is the head gain across pump i (m).

The water flow problem verifies physical consistency by removing the cost minimization objective (25a) while maintaining feasibility constraints (25b)–(25c). These operations represent typical SCADA functionalities in modern WDNs, where operators rely on automated systems for control decisions and hydraulic validation. By manipulating sensor measurements while maintaining apparent physical consistency, the proposed attacks can induce suboptimal pump schedules or false feasibility assessments. Next, we evaluate these impacts through case studies on standard benchmark networks.

5. Case Studies

This section evaluates the proposed strategies on two standard EPANET benchmark networks: Net1 and Net3 [37]. The schematic layouts of these networks are shown in Fig. 3. All system parameters are extracted using the EPANET toolbox on MATLAB (R2024a), with optimization solved using Gurobi 11.0.2. Intrusion detection employs a vectorized CUSUM statistic, with parameters tuned using historical data during normal operation. The detection threshold (τ) is set as the mean of residuals plus three times the standard deviation, while the parameter (b) is selected as the mean of residuals plus 0.5 times the standard deviation. For the chi-squared, detector, the threshold is tuned according to Section 2.3.2. The electricity cost is assumed to be \$0.175 per kWh. For Net1, the WLS relies on 5 sensors across the network, providing approximately 24% coverage of all 21 potential measurement points (12 pipe flows and 9 head measurements). The sensor configuration includes 4 flow sensors monitoring pipes 1, 5, 7, and 9, and a pressure sensors at junction 3. Additionally, we assume that the operator measures the water level at the tank, the pump flow rate, and the demand at Junctions. In the case of Net1, Junction 1 is the only demand junction. The specific placement of sensors for Net3 is omitted for brevity. The full implementation, including all attack formulations, simulation framework, and examples, will
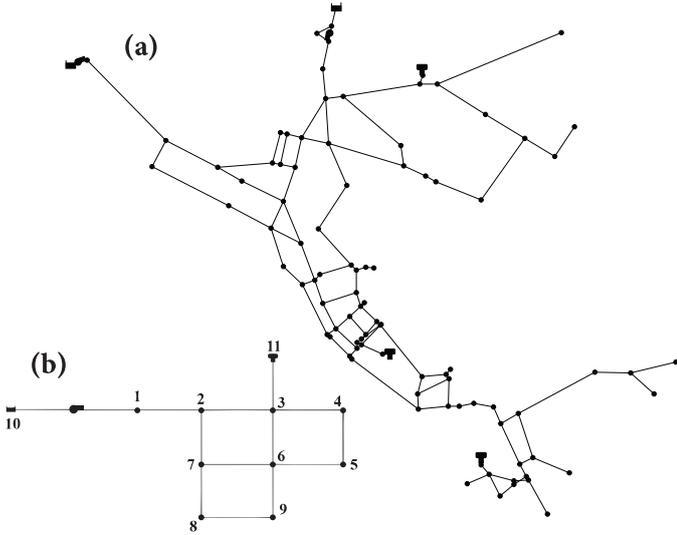
Figure 3: Studied Water Distribution Networks: (a) Net3, (b) Net1

Figure 4: Full-Stealth FDI attack demonstrating stealthy manipulation of pump operations while maintaining detection avoidance

be released on GitHub [3].

Prior to incorporating sensor measurements into SCADA operations (SE, ID, pump control, and water flow) the system performs hydraulic validation checks to ensure physical consistency. These checks verify that received measurements satisfy mass and energy balance constraints across the network. This step is what necessitates that any *successful* sensor manipulation must maintain physical plausibility.

Through these case studies, we investigate: *(i)* How effectively can attacks bypass both detection and physical validation? *(ii)* What impact can they achieve on system operations? *(iii)* Do these approaches remain viable for larger networks?

The following analyses are all conducted on the Net1 network, with Section 5.6 extending to Net3 to demonstrate the scalability of the proposed strategies.

5.1. Full-Stealth FDI Analysis

For the Full-Stealth FDI evaluation, we target three critical measurements: Pump 1 flow sensor, the demand meter at Junction 1, and Pipe 1 flow sensor. The attack magnitude is constrained to 10% of the true measurements. Network observability is ensured through strategic sensor placement: 7 flow sensors across 12 pipes and 6 pressure sensors among 9 junctions, providing 60% coverage of potential measurement points. Simulations are conducted over a 24-hour horizon, with attacks initiated at $t = 11$ and terminated at $t = 20$.

The attack's success depends on its coordinated manipulation of multiple sensors—when the pump flow measurement is altered, corresponding changes are made to connected pipe flows and junction demands to maintain hydraulic consistency. The optimization objective maximizes the magnitude of modifications across accessible measurements while searching for a feasible solution that satisfies both physical and detection constraints. The attacker's ability to construct such a solution critically depends on having access to a sufficient set of hydraulically coupled measurements that can be manipulated simultaneously. This coordination allows the attack to induce suboptimal
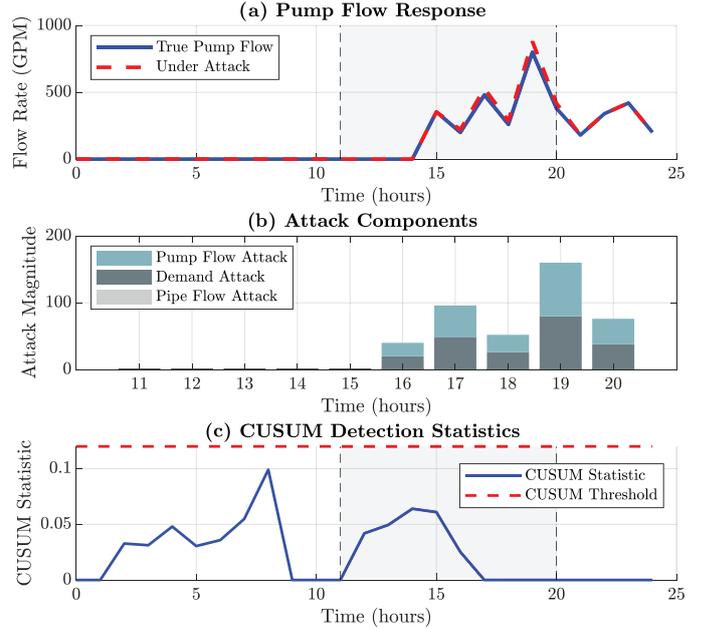
pump operations while ensuring all manipulated measurements appear physically plausible to operators. Fig. 4 demonstrates the impact of the FS-FDI strategy (13) on the optimal pump control problem (25). Between hours 11 and 15, the optimization was unable to find a solution that could increase pump flow or demand while maintaining stealthiness against intrusion detection and physical plausibility constraints. This is likely due to the pump's inactive status during these hours, where any sudden changes would trigger CUSUM detection. However, from hours 15 to 20, the optimization successfully identified opportunities to manipulate both the demand meter at Junction 1 and pump flow sensor, effectively increasing pumping costs while bypassing both physical validation checks and intrusion detection systems.

5.2. Constraint-Relaxed FDI Analysis

Next, we analyze the Hydraulics-Aware FDI strategy under identical timing and targeting parameters. This attack achieved marginally higher impact on pump operations compared to the FS-FDI, suggesting that physical consistency in this scenario posed more restricting constraints than intrusion detection requirements. Fig. 5 demonstrates these effects when maintaining only hydraulic consistency without consideration for detection avoidance.

The relative speed and flow responses in Fig. 5(a,b) show that despite having more degrees of freedom by ignoring detection constraints, the attack's impact remains bounded by the need to satisfy mass and energy balance equations, as evidenced by the preserved head patterns in Fig. 5(c). This relatively constrained impact is also confirmed in Fig. 8, where HA-FDI shows only modest additional cost increases compared to FS-FDI across the full attack duration.

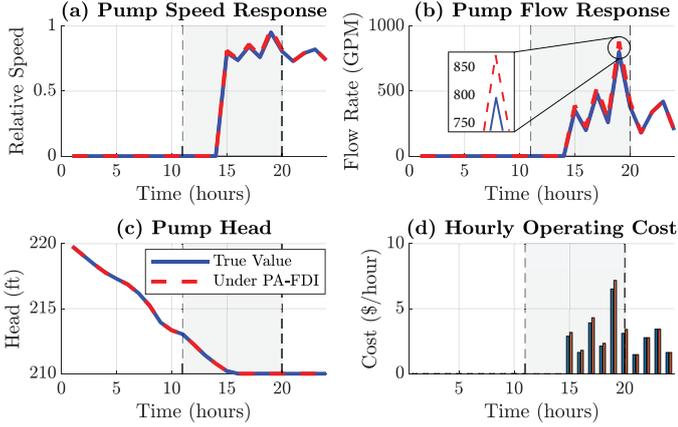Fig. 6 demonstrates how Hydraulics-Unaware attacks that

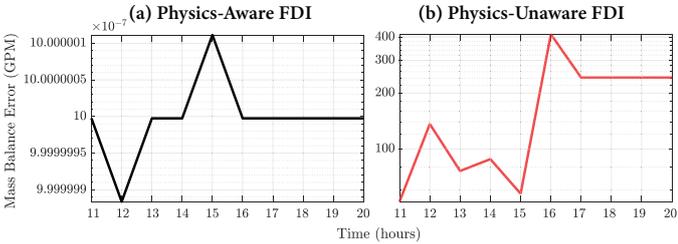Figure 5: Hydraulics-Aware FDI impacts on pump operations and associated costs

Figure 6: Mass Balance Error Validation during: (a) Hydraulics-Aware FDI attack and (b) Hydraulics-Unaware FDI attack

Figure 7: Impact of Random FDI on pump operations and cumulative costs

Figure 8: Comparison of attack strategies' impact on total operational cost

ignore hydraulic relationships lead to obvious physical inconsistencies. While HA-FDI maintains mass conservation with errors below 10^{-7} GPM, HU-FDI causes violations up to 400 GPM. These discrepancies can be immediately flagged by basic physical validation checks, rendering such attacks ineffective in practice. This explains why attackers are likely to prioritize physical constraint satisfaction, even if it limits their ability to maximize operational disruption.

The Random FDI strategy represents a simplistic yet potentially the most common attack scenario, as it requires minimal system knowledge beyond access to measurements. Fig. 7 illustrates its operational impact and shows how the attack forces pump operating points outside optimal regions while remaining within manufacturer-specified curves. A significant increase in the cumulative cost is noticed during the duration of the attack. While these attacks can be easily detected through validation checks and intrusion detection mechanisms, they remain a significant concern for utilities that have not yet implemented such systems. This is particularly concerning as these attacks require minimal sophistication to execute, making them potentially more common in practice.

5.3. Comparison of Attack Strategies

The comparative analysis in Fig. 8 demonstrates the fundamental trade-off between attack impact and stealth.

The results show a clear progression in attack impact: FS-FDI and HA-FDI achieve similar cost increases while maintaining physical plausibility, HU-FDI shows higher impact but fails physical validation, and Random FDI causes the largest
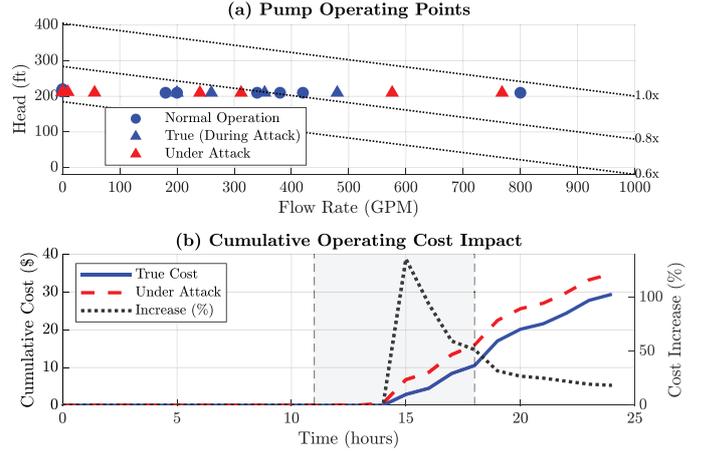
disruption but would be easily detected through both physical and statistical checks. This hierarchy demonstrates how real-world constraints fundamentally limit the ability of FDI attacks to maximize system disruption while maintaining stealthiness, while also highlighting the potential vulnerability of systems lacking detection mechanisms to even simple attacks.

While these results are from a benchmark network, their implications for real-world systems can be significant. In metropolitan water networks where daily pumping costs reach hundreds of thousands of dollars, sustaining a 5% cost increase through stealthy attacks could accumulate to millions in excess operational costs annually as these attacks can persist undetected for extended periods of time.

5.4. Impact on Water Flow Operations

While previous analysis focused on attacks targeting pump control optimization, the proposed strategies are generalizable and can be adapted for other operational objectives.

The attack, executed between $t = 6$ and $t = 14$, manipulates measurements to induce excessive water withdrawal from storage tanks. Fig. 9 demonstrates how a simple Random FDI attack can effectively target water flow management processes. It shows how subtle measurement modifications lead to cumulative tank level reductions. Fig. 9(b) quantifies the total volume impact, an additional 4,874 ft^3 drawn from the tank during the 8-hour attack period. This illustrates how the proposed attack strategies can be generalized beyond cost manipulation, though their effectiveness depends on the attacker's measurement access and the objective they aim to achieve.
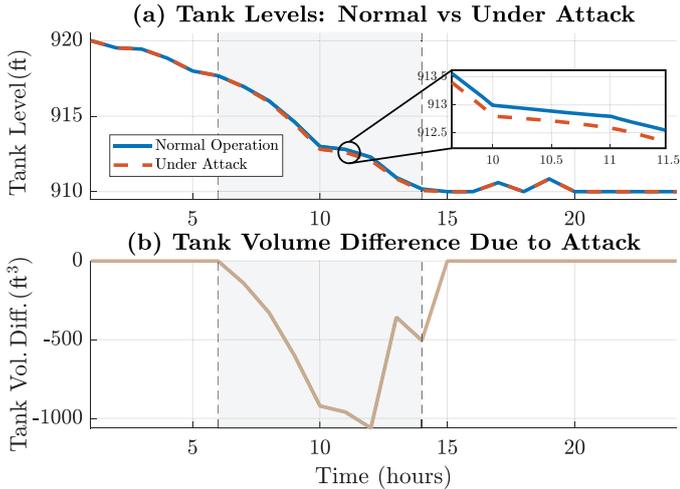
11

Figure 9: Random FDI impact on tank water levels and cumulative volume loss

Figure 10: Comparison of attack impact under Chi-squared and CUSUM detection

The ability to construct effective attacks requires careful selection of target measurements based on both accessibility and hydraulic relationships. While the optimization framework can often find feasible solutions that satisfy physical and detection constraints, achieving specific operational impacts requires strategic targeting of measurements that can influence the desired control variables. However, our simulations demonstrate that even random selection of flow and demand measurements typically yields detrimental effects on system operation, suggesting that the inherent coupling of hydraulic variables makes any measurement manipulation potentially harmful. This highlights the importance of understanding system topology when analyzing vulnerabilities and the need for comprehensive monitoring and detection systems in WDNs.

5.5. Evaluation of Intrusion Detection Methods

To quantify the constraints posed by intrusion detection on attack effectiveness, we compare two commonly employed ID methods described in Section 2.3: a static Chi-squared detector that uses a scalar distance measure to collectively monitor residuals, and a vectorized dynamic CUSUM detection that monitors each measurement independently. Fig. 10 shows the comparative performance of these detectors against an FS-FDI attack targeting pump flow sensor and Junction 1 demand between $t = 13$ and $t = 17$ hours.

The results demonstrate that the Chi-squared detector allows for larger measurement manipulations without triggering alarms as seen by the higher pump flow deviations in the Fig. 10(b). This behavior aligns with previous findings in the literature [1]. The increased attack tolerance under Chi-squared detection stems from its aggregation of residuals, which can mask localized anomalies, while CUSUM's dynamic accumulation of deviations enables better detection of persistent changes in individual measurements. This also suggests that vectorized detection approaches may provide better protection against tailored attacks, though at the cost of increased false alarm rates and more complex implementation due to the need for individual threshold tuning and historical data collection for each measurement point.
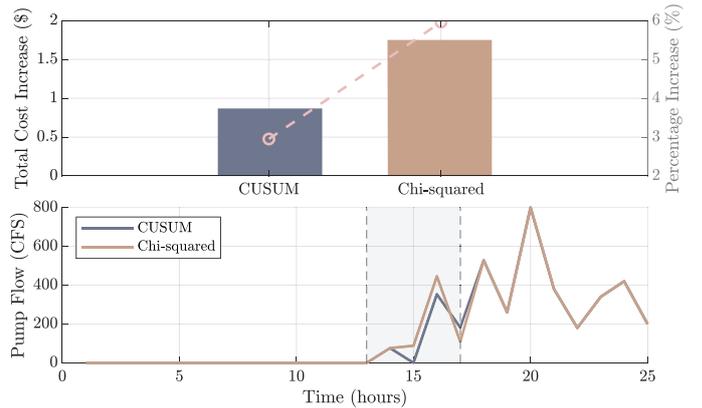
The trade-off between detection sensitivity and false alarms was examined across different threshold values. While tighter thresholds improve attack detection, they lead to increased false alarms during normal operation. Conversely, relaxed thresholds reduce false alarms but create larger blind spots for attackers to exploit. This inherent compromise persists even with carefully tuned parameters, highlighting the fundamental limitations of scalar detection methods against tailored attacks.

5.6. Scalability and Implementation on Larger Networks

The applicability of the proposed attack strategies to larger water distribution networks is primarily constrained by the attacker's access to local system knowledge rather than network size. This is because the designed attacks operate locally—they target specific subsections of the network where the attacker has access to sufficient measurements and system information. When implementing the FS-FDI attack on the larger Net3 network, we observe similar success in manipulating pump operations while maintaining both physical consistency and detection avoidance, despite the network's increased complexity, as seen in Fig. 11. The results demonstrate how small local perturbations can induce significant system-wide impacts in complex networks. During the attack period (hours 10-15), we implemented coordinated modifications to Pump 1's flow rate measurements and demand readings at its adjacent junction, limited to 5% of true values. The impact persisted beyond the attack window, resulting in a 62.74% increase in operational costs over the 24-hour period. This disproportionate impact occurs because the small demand changes force the pump to operate in less efficient regions of its characteristic curve, while the network's interconnected nature means these local inefficiencies cannot be fully compensated by flow from connected pipes.

The scalability of the proposed attacks stems from the fundamentally local nature of hydraulic relationships in water networks. An attacker with knowledge of local topology, hydraulic parameters, and monitoring system configuration can execute these attacks regardless of the overall network size, as long as they can access a sufficient set of hydraulically coupled measurements within their target subsystem. This ensures both practical and computational scalability: the MILP optimization
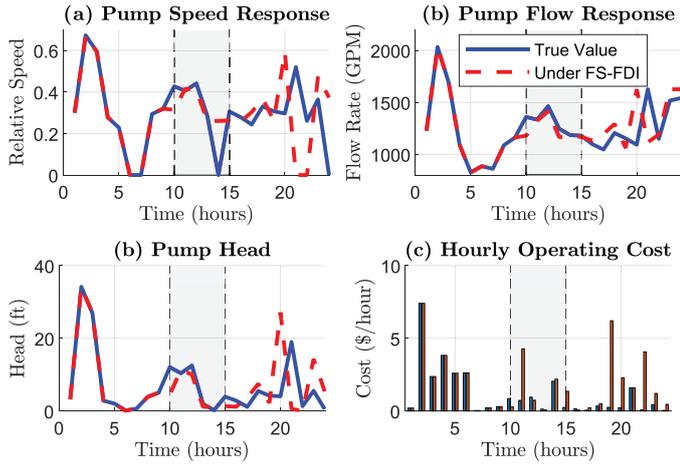
Figure 11: Impact of FS-FDI attack on pump operations in Net3 network

(13) remains tractable as it only needs to consider constraints relevant to the targeted components and their immediate hydraulic connections, independent of the network dimensions.

We note that in our current implementation, the optimizer is allowed to select any pump speed in the interval $[0, s_{max}]$, which can lead to speeds near zero when the solver finds it beneficial from a purely cost-minimization standpoint. In reality, utilities typically enforce practical bounds to maintain pump efficiency and avoid mechanical issues. As our main focus is to demonstrate how stealthy attacks can exploit measurement data to disrupt hydraulic operations (rather than to replicate exact real-world scheduling), we have left such mechanical limits as optional user-defined parameters in the model.

6. Systematic Approach to Attack Design and Execution

This section presents a methodology for designing and implementing stealthy attacks against water distribution networks. We first discuss strategic measurement selection based on network topology, then present a structured attack implementation workflow, and finally examine practical considerations for real-world deployment.

6.1. Strategic Measurement Selection for Attacks

The effectiveness of stealthy attacks depends critically on selecting appropriate measurement targets that satisfy both hydraulic and detection constraints. Our analysis revealed specific network configurations that are particularly vulnerable to manipulation with minimal sensor access:

Four distinct configurations emerged as prime targets for physically-consistent attacks:

- **Boundary Nodes with Demand**: Terminal junctions with a single inflow pipe and demand measurement allow attackers to create physically-consistent false states by manipulating just two measurements. In Net1, Junction 1 exemplifies this vulnerability.
- **Source-Consumer Paths**: Direct hydraulic paths from reservoirs or tanks to demand junctions through pumps or valves. Manipulating flow at the source and demand at the destination creates a physically consistent attack while maximizing energy costs.

- **Storage Elements**: Tanks and their connected pipes represent natural "buffer zones" where flow imbalances can be attributed to level changes. In Net3, manipulating tank levels and adjacent pipe flow readings created consistent false hydraulic states.
- **Flow Distribution Points**: Junctions with multiple connecting pipes require manipulating $n - 1$ flow measurements to maintain mass conservation, making these targets feasible only when sufficient measurement access exists.

The network topology significantly affects vulnerability and attack complexity. The fundamental differences between network structures create distinct attack strategies:

Radial (Tree-like) Networks: Attacks can be highly localized as flow paths are unique between any two points. This topology offers attackers several advantages: *(i)* manipulations remain confined to downstream components, *(ii)* mass conservation requires modifying fewer measurements, and *(iii)* the unique flow paths simplify physical consistency constraints. In Net1, the connection from the reservoir through pump to junction 1 exemplifies this vulnerability.

Looped Networks: The redundant flow paths create hydraulic interdependencies that require more sophisticated approaches: *(i)* Full Loop Manipulation—modify all measurements except one around a complete loop, effectively pushing any physical inconsistency to an unmeasured location; *(ii)* Demand Absorption—target loops containing demand junctions where flow imbalances can be attributed to demand variations; *(iii)* Loop-Breaking—focus on hydraulic control points that enforce directional flow, reducing the problem to a simpler radial-like scenario. Our analysis of Net3 demonstrated how attackers exploiting these properties could maintain physical consistency with minimal sensor manipulations rather than controlling all measurements in a loop.

As illustrated in Fig. 12, both Net3 (a) and Net1 (b) contain multiple potential attack targets (highlighted in red) that require minimal sensor manipulations for the purpose of physically-consistent attacks (such as FS-FDI and HA-FDI attacks). These include terminal demand nodes, pump-reservoir connections, and tank interfaces where attackers can create physically consistent false states with access to just 2-3 measurements.

6.2. Attack Implementation Workflow

To facilitate systematic attack simulation and vulnerability assessment, we developed a comprehensive workflow that guides users from measurement selection to attack execution. This process is formalized in Algorithm 3 (see Appendix B), which outlines the complete procedure for implementing any of the four FDI attack strategies based on available system knowledge and access privileges.

The algorithm provides a structured approach for practitioners to evaluate vulnerabilities without requiring specialized expertise in optimization theory or hydraulic modeling. It includes crucial feasibility checks, constraint validation, and decision points that ensure generated attacks remain within realistic operational parameters. When constraints cannot be satisfied, the framework suggests alternative sensor groupings or parameter adjustments to identify feasible attack scenarios.
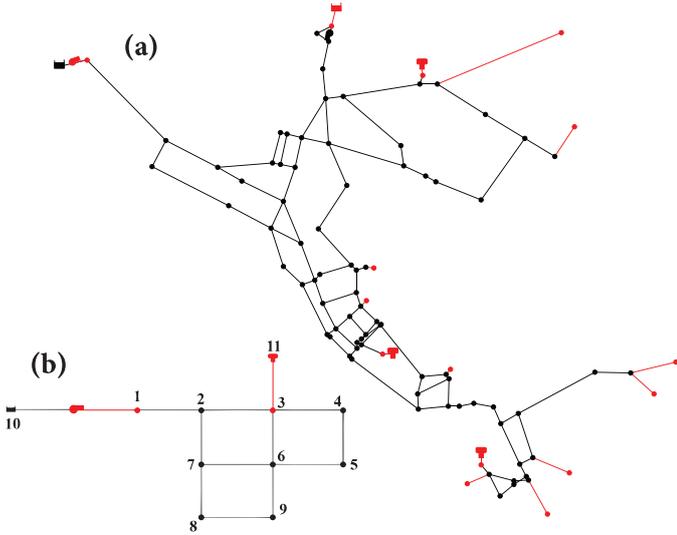
Figure 12: Potential targets for physically-consistent attacks in water distribution networks. Highlighted red components indicate strategic attack points requiring minimal sensor manipulations

To illustrate this workflow, we present a representative attack scenario based on our Net1 case study. The process begins with strategic target selection: the pump flow sensor and Junction 1 demand meter were identified as ideal candidates due to their hydraulic connectivity and operational significance. These components meet the "boundary node with demand" vulnerability pattern described earlier, requiring only two measurement manipulations to maintain physical consistency.

After target selection, the framework performs hydraulic connectivity validation to ensure the selected measurements can be manipulated in a physically consistent manner. For our Net1 example, the mass balance relationship ($Pump\ flow + a_k^p$) − $Pipe\ 1\ flow = (Demand + a_k^d)$ must be preserved. The optimization then computes attack vectors that maximize operational impact while satisfying both physical constraints and detection thresholds.

The attack execution, as demonstrated in Fig. 4, shows how coordinated manipulations during hours 15-20 increased apparent pump flow and demand, inducing a 7-15% cost increase while maintaining physical plausibility and evading detection. This pattern scales effectively to larger networks, as confirmed in our Net3 implementation (Fig. 11), where similar principles generated a 62.74% cost increase using the same algorithmic approach.

This workflow demonstrates that effective attacks do not require comprehensive network knowledge, but rather targeted understanding of critical hydraulic relationships in specific subsystems. The adaptability of our framework allows for systematic vulnerability assessment across diverse network configurations, providing utilities with practical tools to evaluate security weaknesses and develop appropriate countermeasures.

6.3. Implementation Considerations and Practical Aspects

The practical deployment of these attack strategies depends heavily on both the network topology and the attacker's access to system information. Here we discuss key implementation aspects focused on measurement selection and real-world execution constraints.

6.3.1. State Estimation Robustness

The effectiveness of state estimation in real-world WDNs is influenced by measurement noise characteristics, sensor placement density, and model parameter uncertainty. Our implementation accounts for measurement noise through the diagonal weighting matrix W_m, assigning lower weights to sensors with higher variance. While sparse sensor networks—common in water utilities due to installation costs—increase vulnerability to targeted manipulations, they also constrain attackers by limiting the available measurement points for coordinated attacks. Robustness can be improved through strategic redundant sensor placement at hydraulic bottlenecks and critical control points. Additionally, model parameter uncertainty (e.g., in pipe roughness coefficients and demand patterns) introduces natural variability in residuals that detection systems must accommodate, potentially creating margins that sophisticated attackers can exploit while remaining below alarm thresholds.

6.3.2. Implementation and Technical Reproducibility

The implementation of these strategies requires compromising field devices like RTUs and PLCs that interface with physical sensors through industrial protocols [33]. The required system knowledge varies by attack type, while random manipulation may succeed with basic network access, sophisticated attacks require understanding of both system topology and operational patterns, typically gathered through extended network reconnaissance and SCADA traffic monitoring [40].

The attacks are implementable through standard optimization techniques with computational requirements that remain tractable even for large networks. Specifically, the resulting Mixed-Integer Linear Programs (MILPs) scale primarily with the number of targeted sensors and their immediate hydraulic connections rather than with overall network dimensionality. For typical attack configurations targeting 2-3 sensors in a subsystem, the optimization involves approximately $O(10 − 20)$ binary variables for the piecewise linearization. In our computational experiments using Gurobi on standard hardware, these MILPs consistently solve within 0.5-2 seconds even for the larger Net3 network—orders of magnitude faster than typical hydraulic time steps (15-60 minutes) in operational WDNs. This computational efficiency combined with the relatively slow WDN dynamics provides ample time for execution in real-world scenarios. Implementation can be achieved through either hydraulic simulators like EPANET [37] or custom optimization frameworks. While FS-FDI attacks require significant preparation and system knowledge, they represent worst-case scenarios for security evaluation. Given that many utilities lack proper monitoring systems, simpler attacks requiring minimal knowledge may pose more immediate practical threats, highlighting the importance of implementing basic security measures like state estimation and intrusion detection.

6.3.3. Security Recommendations

Based on the analysis of attack strategies and their operational impacts, we propose several recommendations for water utilities. Sensor placement should prioritize hydraulically

coupled measurements to enable cross-validation with critical subsystems like pump stations and storage tanks warranting redundant monitoring through independent sensor types. This enables detection of local inconsistencies while limiting an attacker's ability to coordinate measurement manipulations.

Detection system implementation should prioritize physical validation checks across hydraulically connected components, as these provide fundamental protection regardless of attack *sophistication*. This can be enhanced with statistical monitoring and state estimation when resources permit. From an operational perspective, SCADA system segmentation should account for hydraulic relationships while ensuring control decisions are verified using multiple independent measurements.

7. Paper Limitations and Future Work

While this paper presents a systematic framework for analyzing stealthy false data injection attacks against water distribution networks, several limitations of our current approach should be acknowledged:

- **Deterministic Hydraulic Model.** Our formulation relies primarily on a deterministic hydraulic model with measurement noise as the only source of uncertainty, without fully accounting for epistemic uncertainty in network parameters.
- **Limited State Estimation Methods.** We employ WLS for state estimation, which represents a static estimation approach that does not leverage temporal correlations or dynamic system behavior that more advanced estimators might capture.
- **Localized Attack Scope.** The proposed attack strategies primarily target localized subsystems rather than investigating coordinated attacks against multiple network segments simultaneously.
- **Network Scale Limitations.** The case studies utilize benchmark networks (Net1 and Net3) with limited size and complexity compared to real metropolitan water distribution systems.

These limitations reflect the inherent trade-offs in modeling complex cyber-physical systems, where increased model fidelity must be balanced against computational tractability and analytical clarity. Despite these constraints, our approach successfully demonstrates the vulnerabilities present in water distribution networks and establishes a foundation for more comprehensive security assessment.

Future research should extend this work in several directions: incorporating demand and parameter uncertainties through robust or chance-constrained optimization; implementing advanced state estimation techniques such as Extended Kalman Filters for time-series-based analysis; developing specialized detection algorithms that exploit WDN hydraulic constraints; exploring the integration of distributed observer-based approaches from control theory into water system security frameworks; and experimental validation on larger, operational networks. Particularly promising is the potential application of distributed observer schemes and leader-following control paradigms from the control-theoretic literature to WDN security, which could enable more resilient monitoring architectures while maintaining compatibility with industry-standard tools

like EPANET. These extensions would transform vulnerability analysis into practical security improvements for critical water infrastructure, addressing the multi-faceted challenges in protecting these essential systems against tailored and stealthy cyberattacks.

8. Summary and Concluding Remarks

This paper presents a systematic analysis of stealthy false data injection attacks against water distribution networks. We propose several attack strategies that can bypass both physical validation and intrusion detection while disrupting critical operations like pump control and water flow management. We demonstrate through simulations how attackers with varying levels of system knowledge can manipulate sensor measurements to degrade operational efficiency and hydraulic performance while maintaining apparent physical consistency.

The analysis reveals how attackers can leverage local hydraulic relationships to construct feasible attacks without requiring complete network knowledge. This locality principle ensures attack viability even in large networks, while also highlighting vulnerable subsystems that merit enhanced monitoring. The current lack of basic validation mechanisms in many water utilities means even straightforward sensor attacks can cause significant operational disruptions, emphasizing the urgent need for implementing fundamental monitoring systems.

Looking forward, as water utilities transition toward increased automation and connectivity, the frameworks developed here will become increasingly relevant for systematic vulnerability assessment and security design. Future work should focus on developing resilient control strategies and distributed detection methods that leverage spatial and temporal correlations between measurements, while accounting for constraints like model uncertainties and sensor calibration errors.

References

[1] Ahmed, C.M., Murguia, C., Ruths, J., 2017a. Model-based Attack Detection Scheme for Smart Water Distribution Networks, in: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, ACM, Abu Dhabi United Arab Emirates. pp. 101–113. doi:10.1145/3052973.3053011.

[2] Ahmed, C.M., Palleti, V.R., Mathur, A.P., 2017b. WADI: A water distribution testbed for research in the design of secure cyber physical systems, in: Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks, ACM, Pittsburgh Pennsylvania. pp. 25–28. doi:10.1145/3055366.3055375.

[3] Albustami, A., Taha, A.F., 2024. Fdi-wdns: False data injection attacks on water distribution networks. https://github.com/abdallahbustami/FDI-WDNs. URL: https://github.com/abdallahbustami/FDI-WDNs. accessed: April 25, 2025.

[4] Almalki, S.A., Sheldon, F.T., 2021. Deep learning to improve false data injection attack detection in cooperative intelligent transportation systems, in: 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), IEEE. pp. 1016–1021.

[5] Almalki, S.A., Song, J., 2020. A review on data falsification-based attacks in cooperative intelligent transportation systems. Int. J. Comput. Sci. Secur.(IJCSS) 14, 22.

[6] Andersen, J.H., Powell, R.S., 2000. Implicit state-estimation technique for water network monitoring. Urban Water 2, 123–130. URL: https://linkinghub.elsevier.com/retrieve/pii/S1462075800000509, doi:10.1016/S1462-0758(00)00050-9.

[7] Aoufi, S., Derhab, A., Guerroumi, M., 2020. Survey of false data injection in smart power grid: Attacks, countermeasures

and challenges. Journal of Information Security and Applications 54, 102518. URL: https://www.sciencedirect.com/science/article/pii/S2214212619310713, doi:https://doi.org/10.1016/j.jisa.2020.102518.

[8] Arsene, C.T., Gabrys, B., 2014. Mixed simulation-state estimation of water distribution systems based on a least squares loop flows state estimator. Applied Mathematical Modelling 38, 599–619. URL: https://www.sciencedirect.com/science/article/pii/S0307904X13004149, doi:https://doi.org/10.1016/j.apm.2013.06.012.

[9] Aslam, M.M., Tufail, A., Kim, K.H., Apong, R.A.A.H.M., Raza, M.T., 2023. A Comprehensive Study on Cyber Attacks in Communication Networks in Water Purification and Distribution Plants: Challenges, Vulnerabilities, and Future Prospects. Sensors 23, 7999. doi:10.3390/s23187999.

[10] Bargiela, A., Hainsworth, G.D., 1989. Pressure and Flow Uncertainty in Water Systems. Journal of Water Resources Planning and Management 115, 212–229. URL: https://ascelibrary.org/doi/10.1061/%28ASCE%290733-9496%281989%29115%3A2%28212%29, doi:10.1061/(ASCE)0733-9496(1989)115:2(212).

[11] Bartos, M., Thomas, M., Kim, M.G., Frankel, M., Sela, L., 2024. Online state estimation in water distribution systems via extended kalman filtering. Water Research 264, 122201. URL: https://www.sciencedirect.com/science/article/pii/S004313542401100X, doi:https://doi.org/10.1016/j.watres.2024.122201.

[12] Brentan, B., Rezende, P., Barros, D., Meirelles, G., Luvizotto, E., Izquierdo, J., 2021. Cyber-Attack Detection in Water Distribution Systems Based on Blind Sources Separation Technique. Water 13, 795. doi:10.3390/w13060795.

[13] Brumback, B., Srinath, M., 1987. A chi-square test for fault-detection in kalman filters. IEEE Transactions on Automatic Control 32, 552–554. doi:10.1109/TAC.1987.1104658.

[14] Carrega, J., Lynch, C., Vera, A., 2021. Someone tried to poison a Florida city by hacking into the water treatment system, sheriff says. https://www.cnn.com/2021/02/08/us/oldsmar-florida-hack-water-poison/index.html.

[15] Douglas, H.C., Taormina, R., Galelli, S., 2019. Pressure-Driven Modeling of Cyber-Physical Attacks on Water Distribution Systems. Journal of Water Resources Planning and Management 145, 06019001. doi:10.1061/(ASCE)WR.1943-5452.0001038.

[16] Fayzul, M., Pasha, K., 2018. Development of an Effective Hybrid Method to Detect Cyber-Physical Attack on Water Distribution Systems, in: World Environmental and Water Resources Congress 2018, American Society of Civil Engineers, Minneapolis, Minnesota. pp. 410–421. doi:10.1061/9780784481424.043.

[17] Hassanzadeh, A., Rasekh, A., Galelli, S., Aghashahi, M., Taormina, R., Ostfeld, A., Banks, M.K., 2020. A Review of Cybersecurity Incidents in the Water Sector. Journal of Environmental Engineering 146, 03120003. doi:10.1061/(ASCE)EE.1943-7870.0001686.

[18] Housh, M., Ohar, Z., 2018. Model-based approach for cyber-physical attack detection in water distribution systems. Water Research 139, 132–143. doi:10.1016/j.watres.2018.03.039.

[19] Liang, G., Zhao, J., Luo, F., Weller, S.R., Dong, Z.Y., 2017. A review of false data injection attacks against modern power systems. IEEE Transactions on Smart Grid 8, 1630–1638. doi:10.1109/TSG.2015.2495133.

[20] Linsley, R.K., Franzini, J.B., Freyberg, D.L., Tchobanoglous, G., 1991. Water Resources Engineering. 4th edition ed., McGraw-Hill Science/Engineering/Math, New York.

[21] Liu, Y., Ning, P., Reiter, M.K., 2011. False data injection attacks against state estimation in electric power grids. ACM Trans. Inf. Syst. Secur. 14. URL: https://doi.org/10.1145/1952982.1952995, doi:10.1145/1952982.1952995.

[22] Lyngaas, S., 2024. Russia-linked hacking group suspected of carrying out cyberattack on texas water facility, cybersecurity firm says. URL: https://www.cnn.com/2024/04/17/politics/russia-hacking-group-suspected-texas-water-cyberattack/index.html.

[23] Menke, R., Abraham, E., Parpas, P., Stoianov, I., 2015. Approximation of system components for pump scheduling optimisation. Procedia Engineering 119, 1059–1068. URL: https://www.sciencedirect.com/science/article/pii/S1877705815026053, doi:https://doi.org/10.1016/j.proeng.2015.08.935. computing and Control for the Water Industry (CCWI2015) Sharing the best practice in water management.

[24] Moazeni, F., Khazaei, J., 2021a. Formulating false data injection cyber-attacks on pumps' flow rate resulting in cascading failures in smart water systems. Sustainable Cities and Society 75, 103370. doi:10.1016/j.scs.2021.103370.

[25] Moazeni, F., Khazaei, J., 2021b. Sequential false data injection cyberattacks in water distribution systems targeting storage tanks; a bi-level optimization model. Sustainable Cities and Society 70, 102895. doi:10.1016/j.scs.2021.102895.

[26] Moazeni, F., Khazaei, J., 2022. Detection of Random False Data Injection Cyberattacks in Smart Water Systems Using Optimized Deep Neural Networks. Energies 15, 4832. doi:10.3390/en15134832.

[27] Moazeni, F., Khazaei, J., Mitra, P., 2020. An integrated state-estimation framework for interdependent water and energy systems. Journal of Hydrology 590, 125393. doi:10.1016/j.jhydrol.2020.125393.

[28] Murguia, C., Ruths, J., 2016. CUSUM and Chi-squared attack detection of compromised sensors. 2016 IEEE Conference on Control Applications, CCA 2016 , 474–480doi:10.1109/CCA.2016.7587875.

[29] Page, E.S., 1954. Continuous inspection schemes. Biometrika 41, 100–115.

[30] Pang, Z.H., Fan, L.Z., Dong, Z., Han, Q.L., Liu, G.P., 2022. False data injection attacks against partial sensor measurements of networked control systems. IEEE Transactions on Circuits and Systems II: Express Briefs 69, 149–153. doi:10.1109/TCSII.2021.3073724.

[31] Powell, R.S., Nagar, A.K., Andersen, J., 2000. A Review of Techniques of State-Estimation for On-Line Monitoring of Water Distribution Systems, in: Building Partnerships, American Society of Civil Engineers, Minneapolis, Minnesota, United States. pp. 1–10. URL: http://ascelibrary.org/doi/10.1061/40517%282000%29212, doi:10.1061/40517(2000)212.

[32] Quinonez, R., Giraldo, J., Salazar, L., Bauman, E., Cardenas, A., Lin, Z., 2020. SAVIOR: Securing autonomous vehicles with robust physical invariants. Proceedings of the 29th USENIX Security Symposium , 895–912.

[33] Rajabpour, N., Sedaghat, Y., 2015. A hybrid-based error detection technique for plc-based industrial control systems, in: 2015 IEEE 20th Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–7. doi:10.1109/ETFA.2015.7301525.

[34] Ramotsoela, D.T., Hancke, G.P., Abu-Mahfouz, A.M., 2019. Attack detection in water distribution systems using machine learning. Human-centric Computing and Information Sciences 9, 13. doi:10.1186/s13673-019-0175-8.

[35] Raza, N., Moazeni, F., 2024. Chance-constrained vulnerability assessment of smart water distribution systems against stealthy false data injection attacks. International Journal of Critical Infrastructure Protection 44, 100645. doi:10.1016/j.ijcip.2023.100645.

[36] Romero-Ben, L., Irofti, P., Stoican, F., Puig, V., 2024. Nodal hydraulic head estimation through unscented kalman filter for data-driven leak localization in water networks. IFAC-PapersOnLine 58, 67–72.

[37] Rossman, L.A., Woo, H., Tryby, M., Shang, F., Janke, R., Haxton, T., 2020. Epanet 2.2 user manual; water infrastructure division. Center for Environmental Solutions and Emergency Response .

[38] Taormina, R., Galelli, S., Douglas, H., Tippenhauer, N., Salomons, E., Ostfeld, A., 2019. A toolbox for assessing the impacts of cyber-physical attacks on water distribution systems. Environmental Modelling & Software 112, 46–51. doi:10.1016/j.envsoft.2018.11.008.

[39] Taormina, R., Galelli, S., Tippenhauer, N.O., Salomons, E., Ostfeld, A., Eliades, D.G., Aghashahi, M., Sundararajan, R., Pourahmadi, M., Banks, M.K., Brentan, B.M., Campbell, E., Lima, G., Manzi, D., Ayala-Cabrera, D., Herrera, M., Montalvo, I., Izquierdo, J., Luvizotto, E., Chandy, S.E., Rasekh, A., Barker, Z.A., Campbell, B., Shafiee, M.E., Giacomoni, M., Gatsis, N., Taha, A., Abokifa, A.A., Haddad, K., Lo, C.S., Biswas, P., Pasha, M.F.K., Kc, B., Somasundaram, S.L., Housh, M., Ohar, Z., 2018. Battle of the Attack Detection Algorithms: Disclosing Cyber Attacks on Water Distribution Networks. Journal of Water Resources Planning and Management 144, 04018048. doi:10.1061/(ASCE)WR.1943-5452.0000969.

[40] Torrisi, N., Vukovic, O., Dán, G., Hagdahl, S., 2014. Peekaboo: A gray hole attack on encrypted scada communication using traffic anal-

16

ysis. 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm) , 902–907doi:10.1109/SmartGridComm.2014.7007763.

[41] Tshehla, K.S., Hamam, Y., Abu-Mahfouz, A., 2017a. State estimation in water distribution network: A review. 2017 IEEE 15th International Conference on Industrial Informatics (INDIN) , 1247–1252doi:10.1109/INDIN.2017.8104953.

[42] Tshehla, K.S., Hamam, Y., Abu-Mahfouz, A.M., 2017b. State estimation in water distribution network: A review, in: 2017 IEEE 15th International Conference on Industrial Informatics (INDIN), IEEE, Emden. pp. 1247–1252. URL: http://ieeexplore.ieee.org/document/8104953/, doi:10.1109/INDIN.2017.8104953.

[43] Urbina, D.I., Giraldo, J., Cardenas, A.A., Tippenhauer, N.O., Valente, J., Faisal, M., Ruths, J., Candell, R., Sandberg, H., 2016a. Limiting the impact of stealthy attacks on Industrial Control Systems. Proceedings of the ACM Conference on Computer and Communications Security 24-28-Octo, 1092–1105. doi:10.1145/2976749.2978388.

[44] Urbina, D.I., Giraldo, J.A., Cardenas, A.A., Tippenhauer, N.O., Valente, J., Faisal, M., Ruths, J., Candell, R., Sandberg, H., 2016b. Limiting the Impact of Stealthy Attacks on Industrial Control Systems, in: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, ACM, Vienna Austria. pp. 1092–1105. doi:10.1145/2976749.2978388.

[45] Vrachimis, S.G., Timotheou, S., Eliades, D.G., Polycarpou, M.M., 2019. Iterative Hydraulic Interval State Estimation for Water Distribution Networks. Journal of Water Resources Planning and Management 145, 04018087. URL: https://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0001011, doi:10.1061/(ASCE)WR.1943-5452.0001011.

[46] Yang, Q., Yang, J., Yu, W., An, D., Zhang, N., Zhao, W., 2014. On false data-injection attacks against power system state estimation: Modeling and countermeasures. IEEE Transactions on Parallel and Distributed Systems 25, 717–729. doi:10.1109/TPDS.2013.92.

[47] Zamzam, A.S., Dall'Anese, E., Zhao, C., Taylor, J.A., Sidiropoulos, N.D., 2019. Optimal water–power flow-problem: Formulation and distributed optimal solution. IEEE Transactions on Control of Network Systems 6, 37–47. doi:10.1109/TCNS.2018.2792699.

[48] Zhang, T.Y., Ye, D., Shi, Y., 2023. Decentralized false-data injection attacks against state omniscience: Existence and security analysis. IEEE Transactions on Automatic Control 68, 4634–4649. doi:10.1109/TAC.2022.3209396.

[49] Zhang, T.Y., Ye, D., Yang, G.H., 2024. Ripple effect of cooperative attacks in multi-agent systems: Results on minimum attack targets. Automatica 159, 111307. URL: https://www.sciencedirect.com/science/article/pii/S0005109823004715, doi:https://doi.org/10.1016/j.automatica.2023.111307.

Appendix A. Weighted Least Squared State Estimation

The WLS state estimation minimizes a weighted sum of squared residuals between measured and estimated values, while incorporating hydraulic constraints. The state vector $x \in \mathbb{R}^n$ contains all pipe flows and junction heads:

$$x = \begin{bmatrix} q \\ h \end{bmatrix},$$

where $q \in \mathbb{R}^{n_p}$ represents flows and $h \in \mathbb{R}^{n_j}$ represents junction heads, with $n = n_p + n_j$.

For direct measurements from sensors, we construct a measurement matrix $H_m \in \mathbb{R}^{m \times n}$ where m is the number of sensors:

$$y = H_m x + v,$$

where $y \in \mathbb{R}^m$ is the measurement vector and $v \in \mathbb{R}^m$ is the measurement noise vector.

Measurement noise is modeled through a diagonal weighting matrix W_m, where each diagonal element is the inverse of the corresponding measurement variance:

$$W_m = \text{diag}\left(\frac{1}{\sigma_{\text{flow}}^2}, \ldots, \frac{1}{\sigma_{\text{head}}^2}\right).$$

In our implementation, $\sigma_{\text{flow}} = 0.05$ and $\sigma_{\text{head}} = 0.1$.

Mass balance equations at each junction are incorporated as additional constraints:

$$\sum_{i \in \mathcal{I}_j} q_i - \sum_{o \in O_j} q_o = d_j.$$

These constraints are represented through a matrix $H_{\text{mass}} \in \mathbb{R}^{n_j \times n}$ and vector $z_{\text{mass}} \in \mathbb{R}^{n_j}$ such that $H_{\text{mass}} x = z_{\text{mass}}$. Energy balance across pipes is integrated through linearized head loss equations: $h_i - h_j = r_p \cdot q_p$.

These constraints form matrix $H_{\text{energy}} \in \mathbb{R}^{n_e \times n}$ and vector $z_{\text{energy}} \in \mathbb{R}^{n_e}$ such that $H_{\text{energy}} x = z_{\text{energy}}$.

The final combined system is:

$$\underbrace{\begin{bmatrix} \sqrt{W_m} H_m \\ H_{\text{mass}} \\ H_{\text{energy}} \end{bmatrix}}_{H} x = \underbrace{\begin{bmatrix} \sqrt{W_m} y \\ z_{\text{mass}} \\ z_{\text{energy}} \end{bmatrix}}_{z}$$

The solution then is: $\hat{x} = (H^T H)^{-1} H^T z$, where H and z represent the combined system matrix and vector. This represents the configuration of the WLS SE used in this work.

Appendix B. Comprehensive SFDIA Implementation Framework

Algorithm 3 provides a complete implementation framework for all four attack strategies presented in this work. It guides users through the entire process from sensor selection to attack execution, including parameter specification and constraint verification. The algorithm includes adaptive validation checks that help identify when attacks are infeasible, suggesting alternative approaches based on the available system knowledge and sensor access.

Algorithm 3 Comprehensive SFDIA Implementation Framework

Require: Network configuration, sensor measurements $\boldsymbol{y}_k$, target set $\mathcal{T}$
Ensure: Modified measurements $\boldsymbol{y}_k^a$ for vulnerability assessment
 1: **Input:** WDN configuration, measurements $\boldsymbol{y}_k$, target set $\mathcal{T}$
 2: **Select:** Attack type (FS-FDI, HA-FDI, HU-FDI, R-FDI)
 3: **Initialize:** $\boldsymbol{a}_k^h = \boldsymbol{0}$, $\boldsymbol{a}_k^f = \boldsymbol{0}$, $\boldsymbol{a}_k^d = \boldsymbol{0}$
 4: **Define:** Attack bounds $\alpha_h, \alpha_f, \alpha_d$ (typically 5-10% of original values)
 5: **if** Attack type = *Full-Stealth FDI (FS-FDI)* **then**
 6:     **Input:** SE configuration (matrix $\boldsymbol{H}$, weights $\boldsymbol{W}$, convergence bound ϵ)
 7:     **Input:** ID parameters (threshold τ, bias b for CUSUM or α for χ^2)
 8:     **Input:** Local hydraulic parameters (pipe connectivity, resistance coefficients)
 9:     **Validate topology:** Verify hydraulic connectivity among sensors in $\mathcal{T}$
10:     **if** Validation fails **then**
11:         **Output:** "Selected sensors cannot satisfy physical constraints"
12:         **Suggest:** Alternative sensor groupings; **Exit**
13:     **end if**
14:     Compute residuals: $\boldsymbol{r}_k = \boldsymbol{y}_k - \boldsymbol{h}(\hat{\boldsymbol{x}}_k)$
15:     Solve optimization in Eq. (13) for $\boldsymbol{a}_k^h, \boldsymbol{a}_k^f, \boldsymbol{a}_k^d$
16:     **if** Optimization infeasible **then**
17:         **Output:** "No feasible attack within given constraints"; **Exit**
18:     **end if**
19: **else if** Attack type = *Hydraulics-Aware FDI (HA-FDI)* **then**
20:     **Input:** Local hydraulic parameters (pipe connectivity, resistance coefficients)
21:     **Validate topology:** Verify hydraulic connectivity among sensors in $\mathcal{T}$
22:     **if** Validation fails **then**
23:         **Output:** "Selected sensors cannot satisfy physical constraints"; **Exit**
24:     **end if**
25:     Solve optimization in Eq. (19) for $\boldsymbol{a}_k^h, \boldsymbol{a}_k^f, \boldsymbol{a}_k^d$
26: **else if** Attack type = *Hydraulics-Unaware FDI (HU-FDI)* **then**
27:     **Input:** SE configuration (matrix $\boldsymbol{H}$, weights $\boldsymbol{W}$)
28:     **Input:** ID parameters (threshold τ, bias b for CUSUM or α for χ^2)
29:     **Select:** Optimization-based or closed-form solution approach
30:     **if** Optimization-based **then**
31:         Compute residuals: $\boldsymbol{r}_k = \boldsymbol{y}_k - \boldsymbol{h}(\hat{\boldsymbol{x}}_k)$
32:         Solve optimization (FS-FDI formulation without physical constraints)
33:     **else**
34:         Determine detector type (Vectorized/Scalar CUSUM, Chi-squared)
35:         Generate attack using corresponding equation (16, 17, or 18)
36:     **end if**
37: **else if** Attack type = *Random FDI (R-FDI)* **then**
38:     **Input:** Parameters ($\sigma_d, \sigma_n, \alpha_n, \alpha_s, p_s, \alpha_{max}$)
39:     Initialize drift component: $\boldsymbol{a}_0^{\text{drift}} = \boldsymbol{0}$
40:     **for** each sensor $i \in \mathcal{T}$ **do**
41:         Generate drift, noise, and spike components according to Eq. (20-24)
42:         Combine components and apply magnitude bounds
43:     **end for**
44: **end if**
45: **Apply attack:** $\boldsymbol{y}_k^a = \boldsymbol{y}_k + \boldsymbol{a}_k$
46: **Evaluate:** Process $\boldsymbol{y}_k^a$ through hydraulic solver to assess impact
47: **Return:** Modified measurements and projected operational effects

---