Water Resources Research

RESEARCH ARTICLE
10.1029/2023WR034827

Flood Risks of Cyber-Physical Attacks in a Smart Storm Water System

Chung-Yi Lin[1], **Yi-Chen Ethan Yang**[1], **and Faegheh Moazeni**[1]

[1]Department of Civil and Environmental Engineering, Lehigh University, Bethlehem, PA, USA

Key Points:

- We proposed a mathematical framework for evaluating flood risks of cyber-physical attacks in a smart stormwater system
- False data injection can maliciously increase inflow and reduce the outflow of a targeted detention pond in a smart stormwater system
- Additional flood risks caused by false data injection are higher with smaller, more frequent storms

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y.-C. E. Yang,
yey217@lehigh.edu

Citation:

Lin, C.-Y., Yang, Y.-C. E., & Moazeni, F. (2024). Flood risks of cyber-physical attacks in a smart storm water system. *Water Resources Research*, *60*, e2023WR034827. https://doi.org/10.1029/2023WR034827

Received 4 MAR 2023
Accepted 16 DEC 2023

Abstract The rise in smart water technologies has introduced new cybersecurity vulnerabilities for water infrastructures. However, the implications of cyber-physical attacks on the systems like urban drainage systems remain underexplored. This research delves into this gap, introducing a method to quantify flood risks in the face of cyber-physical threats. We apply this approach to a smart stormwater system—a real-time controlled network of pond-conduit configurations, fitted with water level detectors and gate regulators. Our focus is on a specific cyber-physical threat: false data injection (FDI). In FDI attacks, adversaries introduce deceptive data that mimics legitimate system noises, evading detection. Our risk assessment incorporates factors like sensor noises and weather prediction uncertainties. Findings reveal that FDIs can amplify flood risks by feeding the control system false data, leading to erroneous outflow directives. Notably, FDI attacks can reshape flood risk dynamics across different storm intensities, accentuating flood risks during less severe but more frequent storms. This study offers valuable insights for strategizing investments in smart stormwater systems, keeping cyber-physical threats in perspective. Furthermore, our risk quantification method can be extended to other water system networks, such as irrigation channels and multi-reservoir systems, aiding in cyber-defense planning.

1. Introduction

Cyber-physical systems (CPS) interconnect physical infrastructures with digital networks, often leveraging smart water technologies. These technologies enable automated monitoring and operation, facilitating remote real-time control of CPS. However, they also present new cybersecurity challenges to contemporary water systems (Tuptuk et al., 2021). For example, in 2013, the Bowman Avenue Dam, located 30 miles north of Manhattan, was the target of a cyber-attack (Hassanzadeh et al., 2020). Similarly, in 2021, the control system of a water treatment plant in Oldsmar, Florida, was also hacked, enabling the attacker to remotely manipulate the levels of sodium hydroxide in the plant's water supply (Bergal, 2021). Recognizing these threats, the US Federal government recently released a water sector action plan to expand public-private cybersecurity partnerships (The White House, 2022), underscoring the growing concerns over cybersecurity in the water industry.

In managing urban stormwater systems, effective strategies are essential to address issues like flooding, pollution, and erosion (Burian & Edwards, 2002; Fletcher et al., 2015; Jongman, 2018; Ministry of the Environment, 2003). Traditional methods often utilize nature-based solutions such as detention ponds with gravity-driven passive control to manage runoffs and diminish peak outflows (Huang et al., 2020; Van Meter et al., 2011). This involves optimizing properties such as pond capacities, conduit dimensions, and relative invert elevations of components (Froise & Burges, 1978; Yeh & Labadie, 1997). However, passive systems, where factors like outflow gate openings remain static, can be challenging and expensive to modify in response to changing weather patterns or land use once established (Shishegar et al., 2018). As a solution, smart water technologies, including sensors, actuators, and cloud storage, have been introduced to enhance flexibility and efficiency through real-time control (RTC) while being cost-effective (Gaborit et al., 2013; Mullapudi et al., 2017; Piro et al., 2019). For example, controlling upstream pond outflows proactively based on real-time water level data and weather forecasts can help regulate peak flow at a stormwater system's outlet (Sadler et al., 2020; Shishegar et al., 2021; Wong & Kerkez, 2018). Such systems integrating smart water technologies are termed smart stormwater systems (Bartos et al., 2018).

Smart stormwater systems offer notable advantages for flood control, yet their inherent susceptibility to cyber-physical threats cannot be overlooked (Kriaa et al., 2015; Shin et al., 2020). The foundational technology of these systems, termed Cyber Physical Systems (CPS), typically relies on Supervisory Control and Data

Acquisition (SCADA) systems that encompass both regulatory and supervisory layers (Amin et al., 2013). The supervisory layer gathers sensor data and estimates states, such as water levels, for fault detection. The processed data is then distributed to Programmable Logic Controllers (PLCs) via communication channels. Within the regulatory layer, these PLCs deduce control signals like ideal gate positions and convey them to, for instance, gate actuators, through localized networks. The entire communication process within these control systems is prone to cyber-physical attacks. Generally, these attacks fall under two categories: denial-of-service (DoS) attacks, which aim to disable the communication network by overwhelming the system with random data, and deception attacks, which misguide by modifying control objectives (i.e., setpoints) or feeding false sensor readings. Considering that the core design of most CPS involves a feedback control mechanism dependent on sensor data, even minor deceptive actions can have significant real-world consequences.

Elaborating further, the security dynamics in smart stormwater systems are distinctively different from those in traditional information technology (IT). In IT, the primary concern is often binary: whether an attack successfully breaches the SCADA system or not. However, in the context of smart stormwater systems, the emphasis shifts to the real-world implications of such intrusion, specifically, potential flood events. In particular, we focus on a type of deception attack known as "false data injection" (Mo & Sinopoli, 2010). This type of attack underscores the intricate interplay between cyber intrusions and their tangible impacts within a smart stormwater framework. This study aims to bridge the understanding between cyber vulnerabilities and their direct physical outcomes, especially in the realm of smart stormwater management.

Stealthy false data injection (FDI) refers to the act of injecting false sensor measurements into the SCADA system without being detected, ultimately compromising the system's operations (Moazeni & Khazaei, 2021). The scope of FDI encompasses the corruption of data at the sensor level or during the transmission phase between sensors and the SCADA system (Taormina et al., 2018). Successful intrusions often manipulate the system's inherent allowance for minor sensor discrepancies, known as sensor noises. Thus, the injected data essentially masquerades as altered sensor noises. Such contaminated measurements can bias control decisions, such as regulating pond outflows. The ramifications of these biased decisions can extend to tangible system damages like flooding, leading to the term "cyber-physical attacks."

The phenomenon of FDI has been previously analyzed in contexts like water distribution systems (Ahmed et al., 2017; Moazeni & Khazaei, 2021), irrigation canals (Amin et al., 2012, 2013), and water treatment plants (Kumar et al., 2021). However, its impacts have not yet been explored in a smart stormwater system. The flooding impacts of FDI are intricately linked not just to the system's vulnerabilities, but also to the intensity and patterns of prevailing natural storm events.

Additionally, while there's a growing body of research addressing the implications of false data injection (FDI) attacks in CPS, the domain of risk assessment for these attacks remains under-explored. One notable exception is the work of Depoy et al. (2005), who proposed a high-level risk assessment framework for CPS in large-scale critical infrastructures such as water distribution systems. This framework considers both physical and cyber security threats and allows decision-makers to label each threat component as high, medium, or low based on their expertise. Later, the framework was developed into a computer program to assist with cyber-physical systems risk assessment (DePoy et al., 2006). However, a specific mathematical framework for quantifying flood risks under the cyber-physical attack of FDI is still lacking.

Therefore, this study aims to analyze FDI impacts and develop a flood risks quantification method for FDI in a smart stormwater system. We operate under the assumption that FDI attacks are carried out by stealthy attackers who target a specific objective (i.e., to flood a pond) and possess complete knowledge of the CPS (i.e., know how to attack). Namely, this approach represents a worst-case scenario evaluation since we also assume that FDI attacks 100% occur. The study is divided into three tasks: (a) developing a mathematical framework to link FDI and physical responses, (b) examining the FDI impacts in a smart stormwater system, and (c) quantifying the additional flood risks caused by FDI and associated sensor noises and weather forecast uncertainties. As smart stormwater systems are not widely used in practice, this study uses a hypothetical case based on the real-world pond network layout to initiate the discussion of flood risks under FDI and to aid future system development, including the development of defense strategies.

The article is structured as follows. In Section 2, the methods used to accomplish the three aforementioned tasks are presented. Section 3 introduces the study area and the experiment setup. Results are shown in Section 4.

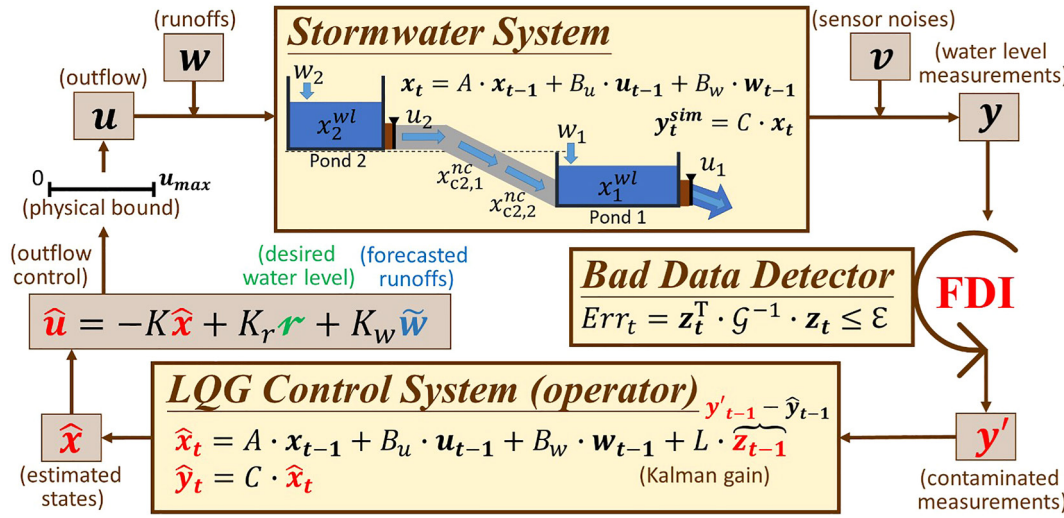

Figure 1. A mathematical framework for smart stormwater systems consists of a stormwater (drainage) system representative, a LQG control system (the operator of the system), and a bad data detector. The red-colored variables indicate the polluted path to the outflow control ($\hat{\boldsymbol{u}}$), where $\boldsymbol{y}'$ is the contaminated measurements from FDI.

Discussion of defending strategies and model limitations is shown in Section 5, followed by the Conclusions in Section 6.

2. Methods

2.1. Mathematical Framework for Smart Stormwater Systems

This study presents a mathematical framework (Figure 1) that abstracts a smart stormwater system into three main components: (a) a stormwater system representative, (b) an outflow control system, and (c) a bad data detector. The stormwater system is modeled using a linear state-space model that simulates the water balance and water level dynamics in a pond-conduit network. A time-invariant linear quadratic gaussian (LQG) controller is employed to control the system and assimilate water level measurements into model estimates using the Kalman filter to regulate outflow controls. A χ^2 detector (Mo & Sinopoli, 2010) is adopted to be a bad data detector. The detector will ensure the eligibility of the sensor measurements before passing them to the LQG controller. These three components comprise a feedback control loop. The study elaborates on these components in the following sections.

2.1.1. State-Space Representation for the Pond-Conduit Network

In our study, we considered a stormwater system consisting of a network of ponds and conduits (Figure 1). This system can be further simplified into a node-link structure, where the nodes represent storages (i.e., ponds) and the links represent the network topology (i.e., conduits). The state-space model (Equation 1) has been successfully adopted to simulate the water balance dynamics in such network-based systems (Ahmed et al., 2017; Schuurmans, 1997; Wong & Kerkez, 2018).

$$\boldsymbol{x}_t = A \cdot \boldsymbol{x}_{t-1} + B_u \cdot \boldsymbol{u}_{t-1} + B_w \cdot \boldsymbol{w}_{t-1} \tag{1}$$

$$\boldsymbol{y}_t^{sim} = C \cdot \boldsymbol{x}_t \tag{2}$$

where $\boldsymbol{x}_t$ [cm] is a vector of states, including the water level of ponds and the flow quantities associated with each conduit segment. The flow quantities are represented by the water level change of the source pond. The inputs term $\boldsymbol{u}_{t-1}$ [cm] is a vector of pond outflows represented by the water level change of the source pond, and $\boldsymbol{w}_{t-1}$ [cm] is a vector of runoffs represented by the water level change of the destination pond. The subscript $t \in \mathcal{T} = \{1, 2, \ldots, T\}$ denote the time step in a discrete simulation system, where $\mathcal{T}$ is a set of simulation time steps, and T is the number of time steps in a simulation. The coefficients A, B_u, and B_w are state, control, and disturbance matrixes, respectively. These matrices represent the dynamic behavior of states, control, and disturbances as time approaches infinity. Matrix A is a square matrix with a dimension equal to the number of states in $\boldsymbol{x}_t$. The diagonal elements corresponding to water level states have the value one; otherwise, zero. The off-diagonal

non-zero elements represent the routing process in conduits. Those non-zero elements have the value $\frac{a_{s,\text{source}}}{n_c \cdot a_{s,\text{destination}}}$, where n_c is the number of segments in a conduct c and $\frac{a_{s,\text{source}}}{a_{s,\text{destination}}}$ is the ratio of the source pond area and the destination pond area. Matrix B_u has the dimension of the number of states in $\boldsymbol{x}_t$ times the number of ponds. In the matrix B_u, pond-to-conduit-inlet elements have the value of one, and conduit-outlet-to-pond elements have the value of negative one; otherwise, zero. Matrix B_w has the same dimension as B_u. The only non-zero elements are those linking sub-catchment runoff to a pond, which have the value of one. We expand Equation 1 and illustrate the water level dynamics of Pond 1 in a two-pond system (Figure 1) in Equation 3.

$$x^{\text{wl}}_{1,t} = x^{\text{wl}}_{1,t-1} + \overbrace{\underbrace{\frac{a_{s,2}}{n_{c2} a_{s,1}}\left(x^{\text{nc}}_{c2,1,t-1} + x^{\text{nc}}_{c2,2,t-1}\right)}_{\text{Conduit inflow from Pond 2}} + \underbrace{u_{1,t-1}}_{\text{Outflow}} + \underbrace{w_{1,t-1}}_{\text{Runoff}}}^{\text{Water level change of Pond 1 from different sources}} \tag{3}$$

where x^{wl} is pond water level states, and $x^{\text{nc}}_{c,s}$ is the water quantity state of segment s in a conduct c. The dynamic of the water level in a pond is equal to the water level at the previous time step (the first term) plus the water level changes resulting from the conduit inflows of the upstream ponds (the second term), the pond's outflow (the third term), and the runoffs (the fourth term) from its sub-catchments. The output matrix (C) in Equation 2 collects the simulated water level ($\boldsymbol{y}^{\boldsymbol{sim}}_t$ [cm]) information from $\boldsymbol{x}_t$. Matrix C has the dimension of the number of ponds times the number of states in $\boldsymbol{x}_t$. We provide a more thorough example with a three-pond system in Supporting Information S1 (Text S1) to demonstrate the construction of a state-space model from a given pond-conduit network.

The sensor measurements of water levels ($\boldsymbol{y}_t$ [cm]) are expressed in Equation 4.

$$\boldsymbol{y}_t = \Gamma^{\text{h}}_t \cdot \left(\boldsymbol{y}^{\boldsymbol{sim}}_t + \boldsymbol{v}_t\right) + \Gamma^{a}_t \cdot \boldsymbol{y}^{\boldsymbol{a}}_t \tag{4}$$

where Γ^h and Γ^a are indicator matrixes (i.e., 0 or 1) showing the time steps and sensors that are healthy or being attacked (i.e., FDI), respectively. The term $\boldsymbol{v}_t$ [cm] is a vector of the sensor noises, which are assumed to be Gaussian white noises. The term $\boldsymbol{y}^{\boldsymbol{a}}_t$ [cm] is a vector of the injected false data time step t.

2.1.2. LQG Controller for System Operation

The second component of the proposed framework is the LQG controller, which is composed of two main parts: the linear quadratic estimator (LQE) and the linear quadratic regulator (LQR). The LQE is the state observer that estimates future states ($\hat{\boldsymbol{x}}_t$ [cm]) by assimilating model predictions ($\hat{\boldsymbol{y}}$ [cm]) and water level measurements through a Kalman filter as shown below:

$$\hat{\boldsymbol{x}}_t = A \cdot \hat{\boldsymbol{x}}_{t-1} + B_u \cdot \boldsymbol{u}_{t-1} + B_w \cdot \boldsymbol{w}_{t-1} + L \cdot \boldsymbol{z}_{t-1} \tag{5}$$

$$\boldsymbol{z}_t = \boldsymbol{y}_t - \hat{\boldsymbol{y}}_t \tag{6}$$

$$\hat{\boldsymbol{y}}_t = C \cdot \hat{\boldsymbol{x}}_t \tag{7}$$

In LQE (Equations 5–7), we adopt the same state-space model (Equations 1 and 2) as the prediction model. The Kalman gain matrix, denoted by L, is used to adjust the current prediction using the differences between predictions and measurements at the previous time step ($\boldsymbol{z}_{t-1}$ [cm]; Equation 6). The value of L is determined based on the level of sensor noises and the uncertainty in weather forecast. When the sensor noises are smaller than the forecast uncertainty, the value of L is closer to 1, indicating that the estimated state ($\hat{\boldsymbol{x}}_t$) relies more on the measurements. Conversely, when the forecast uncertainty is larger than the sensor noises, L is closer to 0, indicating that $\hat{\boldsymbol{x}}_t$ depends more on the model prediction. The calculation for L is presented in (Text S2 of the Supporting Information S1).

LQR is a closed-loop feedback control method (Mo et al., 2010) that will use $\hat{\boldsymbol{x}}_t$ from LQE to determine the optimal control based on an overall objective function (Equation 8; J [cm^2]):

$$J = \hat{\boldsymbol{x}}^{\text{T}}_T Q \hat{\boldsymbol{x}}_T + \sum_{t=1}^{T-1}\left(\hat{\boldsymbol{x}}^{\text{T}}_t Q \hat{\boldsymbol{x}}_t + \hat{\boldsymbol{u}}^{\text{T}}_t R \hat{\boldsymbol{u}}_t\right) \tag{8}$$

where T is the total time step, Q is a weight matrix for the control error (e.g., deviation of desired water level), and R is a weight matrix for the control cost (e.g., power consumption of gate controllers). The term $\hat{\boldsymbol{x}}^{\text{T}}_T Q \hat{\boldsymbol{x}}_T$ is the control error at the terminal step. The analytical solution of this linear optimization problem to minimize J is $\hat{\boldsymbol{u}}_t = -K\hat{\boldsymbol{x}}_t$, where $\hat{\boldsymbol{u}}_t$ is the optimal control at the time step t, and K is the feedback gain matrix solved by the Riccati equation

(Kučera, 1973). In addition to the information derived from the measurements (i.e., $\hat{\boldsymbol{x}}_t$), we adopt future information, such as future desired water levels ($\boldsymbol{r}_t$ [cm]) and forecasted runoffs ($\tilde{\boldsymbol{w}}_t$ [cm]), to adjust the outflow control ($\hat{\boldsymbol{u}}_t$ [cm]):

$$\hat{\boldsymbol{u}}_t = -K \cdot \hat{\boldsymbol{x}}_t + K_r \cdot \boldsymbol{r}_t + K_w \cdot \tilde{\boldsymbol{w}}_t \tag{9}$$

where K_r and K_w are two corresponding feedforward gain matrixes for $\boldsymbol{r}_t$ and $\tilde{\boldsymbol{w}}_t$, respectively. For example, if the system foresees a large incoming runoff or a decrease in desired water levels (i.e., control target), outflow control ($\hat{\boldsymbol{u}}_t$) will be enlarged. We provide the detailed derivation of K, K_r, and K_w in Text S2 of the Supporting Information S1.

The actual controllable outflows (defined as a negative value), however, are limited to the available water in ponds ($\boldsymbol{u}_t^{aw}$ [cm]) and the physical properties of gravity-driven outflows ($\boldsymbol{u}_t^{uc}$ [cm]; assuming no pumps were installed) as shown in Equation 10.

$$\boldsymbol{u}_t = \max\left(\min(0, \hat{\boldsymbol{u}}_t), -\boldsymbol{u}_t^{aw}, -\boldsymbol{u}_t^{uc}\right) \tag{10}$$

$$\boldsymbol{u}_t^{aw} = \boldsymbol{y}_t^{sim} \tag{11}$$

$$\boldsymbol{u}_t^{uc} = \boldsymbol{c}_g \times \boldsymbol{\mu} \times \boldsymbol{a}_g \times \sqrt{2g\boldsymbol{y}_t^{act}} \times \left(\frac{dt}{\boldsymbol{a}_s}\right) \tag{12}$$

$$\boldsymbol{y}_t^{act} = \min\left(\boldsymbol{y}^{max}, max\left(0, \boldsymbol{y}_t^{sim}\right)\right) \tag{13}$$

The available water in ponds ($\boldsymbol{u}_t^{aw}$) is equal to $\boldsymbol{y}_t^{sim}$ (Equation 11). The maximum gravity-driven outflows ($\boldsymbol{u}_t^{uc}$) is computed under the assumption of full pipe flow (Equation 12), where $\boldsymbol{c}_g$ is a calibrated gate coefficient, $\boldsymbol{\mu}$ is the coefficient of contraction (often set to 0.65; Rossman, 2010), $\boldsymbol{a}_g$ is the cross-section area of the maximum gate opening (e.g., orifice), g is gravitational acceleration, and the actual water level in a pond ($\boldsymbol{y}_t^{act}$ [cm]; Equation 13) is bounded between 0 (i.e., no negative water level) and the maximum depth of ponds ($\boldsymbol{y}_t^{max}$ [cm]). The term $\frac{dt}{\boldsymbol{a}_s}$ converts the unit from flow rate to the ponds' water level change, where dt is the simulation time interval and $\boldsymbol{a}_s$ is a vector of the surface area of ponds, a function of $\boldsymbol{y}_t^{act}$.

2.1.3. Bad Data Detector

Bad data is defined as any measurements that deviate from the estimated values' tolerance range. This study adopted a χ^2 detector (Mo et al., 2010) as a bad data detector in Equation 14.

$$\mathrm{Err}_t = \boldsymbol{z}_t^{\mathrm{T}} \cdot \mathcal{G}^{-1} \cdot \boldsymbol{z}_t \leq \varepsilon \tag{14}$$

where $\boldsymbol{z}_t$ is a vector of the water level differences between LQE estimated values and measurements at the time step t as shown in Equation 6, $\mathcal{G}$ is a sensor weight matrix, and ε is an operator-selected threshold for measured errors tolerance (see the next section). We further denote $\boldsymbol{z}_t^T \cdot \mathcal{G}^{-1} \cdot \boldsymbol{z}_t$ as Err_t [cm^2]. The assumption of white noise allows us to consider the error term Err_t as the sum of the weighted squares of independent and identically distributed Gaussian sensor noises (i.e., inter-product of $\boldsymbol{z}_t$). Such that Err_t follows a χ^2 distribution with a degree of freedom equivalent to the number of water level sensors. As sensor noise is a stochastic component of the system, we must design operational rules and attacking strategies from a probabilistic perspective. It is worth noting that $\boldsymbol{u}_{t-1}$, $\boldsymbol{w}_{t-1}$, and $\boldsymbol{z}_{t-1}$ in Equation 5 become known values for the next time step t. The statistic Err_t is independent to forecasted runoffs ($\tilde{\boldsymbol{w}}_t$). Hence, it is not affected by the forecast uncertainty in identifying bad data.

2.2. FDI Impact Evaluation

2.2.1. Operation and FDI Strategies

This study evaluates FDI impact and quantifies the risk associated to FDI cyber-attacks given a set of operation (the operator's perspective) and FDI (the attacker's perspective) strategies. The operation strategy ($\mathcal{O}$) and FDI strategy ($\mathcal{A}$) in a smart stormwater system ($\mathcal{N}$) are defined as:

$$\begin{cases} \mathcal{N} = \{A, B_u, B_w, C, \mathcal{G}, L, K, K_r, K_w\} \\ \mathcal{O} = \{\mathcal{R}, p\}, \\ \mathcal{A} = \{\Gamma^a, p^a\} \end{cases} \tag{15}$$

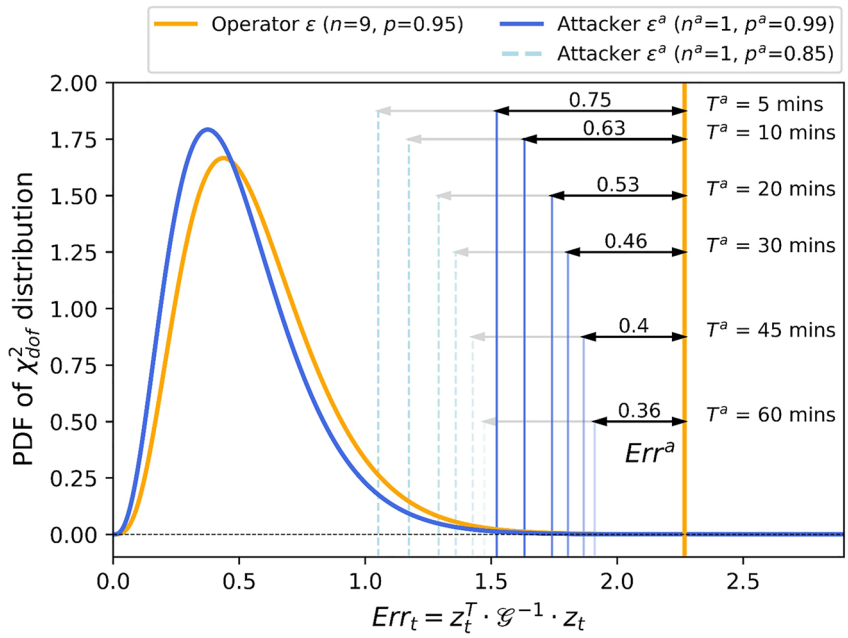

Figure 2. χ^2 distributions for an operator (yellow lines) and attacker (blue lines) to determine corresponding thresholds, ε and ε^a, based on their operation and attack strategy. The attackable range (Err^a) decreases as the number of attacking time steps (T^a) increases and increases as the designed probability of a successful FDI (p^a) decreases (dashed lines).

where p is the designed probability that the detector alarm will not be triggered in a healthy system during a storm event. Namely, $1 - p$ is the probability of a false alarm.

The designed probability of a successful FDI given attacking sensors and time steps (Γ^a) in a storm event is denoted as p^a. Next, we convert p and p^a into the corresponding single-time step thresholds for an operator threshold (ε) and an attacker threshold (ε^a) in Equation 16.

$$\begin{cases} \varepsilon = {\chi^2_{\text{dof}=n}}^{-1}\left(p^{1/T}\right), \\ \varepsilon^a = {\chi^2_{\text{dof}=n-n^a}}^{-1}\left((p^a)^{1/T^a}\right) \end{cases} \tag{16}$$

where T is the total control time steps, n is the number of sensors, T^a is the total attacking time steps, and n^a is the number of attacking sensors. Degree of freedom of a χ^2 distribution is denoted as dof. The decrease in ε^a's dof is because we replace some random sensor noises with a known deterministic false data. The difference between ε and ε^a is defined as an attackable range (Err^a [cm^2]). It represents the maximum magnitude of total injected false data at a single control time step that satisfies the designed attacking strategy. Figure 2 visualizes the concept of ε (yellow lines) and ε^a (blue lines) and indicates Err^a with different T^a (solid lines; dt is equal to 1 min in Figure 2) and different p^a (dashed lines). From the operator's viewpoint, if p is set to a very large value, it increases the targeted space for the attacker (i.e., wider Err^a); however, if p is set to a very low level, then false alarms may lose their significance and get overlooked by operators. From the attacker's viewpoint, if p^a is set too large, decreased Err^a might limit the goal to flood a pond; however, if p^a is set too low, there is a higher chance the attack will be detected and blocked. Being detected might trigger the system upgrade against future attacks, which an attacker will want to avoid.

2.2.2. Solving FDI With Optimization

The proposed mathematical framework can be framed into an optimization model from an attacker viewpoint, which allows us to analyze the maximum FDI impacts that could bring to the system given the operation and attacking strategies (i.e., $\mathcal{O}$ and $\mathcal{A}$, respectively). Specifically, we solve the injected false data ($\boldsymbol{y}_t^a$ [cm]) by formulating the proposed mathematical framework into a deterministic mixed-integer quadratically constrained programming (MIQCP) problem. The stealthy attacker is well-knowledge about the stormwater system ($\mathcal{N}$) and

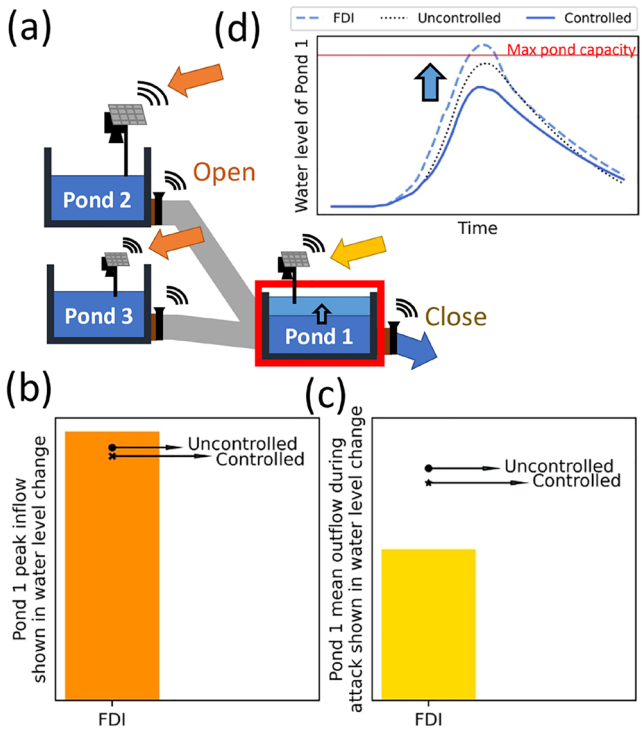

Figure 3. Disclosure of potential FDI consequences with Pond 1 as the flooding target. (a) A smart stormwater system with three ponds. (b) Water level impact of Pond 1 from enlarging inflow caused by FDI. (c) Water level impact of Pond 1 from reducing outflow caused by FDI. (d) The water level time series of Pond 1 under uncontrolled, controlled, and FDI scenarios.

the operation strategy ($\mathcal{O}$) and know how to solve the optimization problem. In addition, we also assume they have a target pond (s_{target}) to flood. Hence, the attacker's objective function in this MIQCP problem is:

$$\text{Obj} = \text{Max}\left\{ \sum y^{\text{sim}}_{s_{\text{target}},t} \right\} \tag{17}$$

which is subjected to Equations 1–13 (except Equations 3 and 8), Equations 18 and 19 with no sensor noises ($\boldsymbol{v}_t = 0$) as $E[\boldsymbol{v}_t] = 0$ and a perfect weather forecast ($\tilde{\boldsymbol{w}}_t = \boldsymbol{w}_t$).

$$\boldsymbol{z}_t^{\mathrm{T}} \cdot \mathcal{G}^{-1} \cdot \boldsymbol{z}_t \le \text{Err}^a \tag{18}$$

$$\boldsymbol{x}_{\mathbf{t}_s} = \hat{\boldsymbol{x}}_{\mathbf{t}_s} = \boldsymbol{x}^h_{\mathbf{t}_s}, \boldsymbol{u}_{\mathbf{t}_s} = \boldsymbol{u}^h_{\mathbf{t}_s}, t \in \mathcal{T}^a = \{t_s + 1, t_s + 2, \ldots, t_s + T^a\} \tag{19}$$

We substitute the bad data detector (Equation 14) with Equation 18 since the only source of $\boldsymbol{z}_t$ is $\boldsymbol{y}^a_t$ in a deterministic setup. The adoption of Err^a is to ensure the successful FDI rate is at least p^a (i.e., rate of not being detected). Since we only need to solve $\boldsymbol{y}^a_t$ for the time steps having FDI, the initial values of this MIQCP problem (i.e., $\boldsymbol{x}_{\mathbf{t}_s}$, $\hat{\boldsymbol{x}}_{\mathbf{t}_s}$, and $\boldsymbol{u}_{\mathbf{t}_s}$) are equal to the values in a healthy system at the time t_s (i.e., $\boldsymbol{x}^h_{\mathbf{t}_s}$ and $\boldsymbol{u}^h_{\mathbf{t}_s}$; Equation 19). We show a complete MIQCP problem in Text S3 of the Supporting Information S1.

2.2.3. Consequences of FDI

To clarify the concept for our audience, we have included an example of a simple three-pond system to demonstrate two potential FDI consequences in Figure 3 before diving into risk quantification.

In this example, Pond 1 is the target of the attack. The first type of FDI impact is when the attacker can manipulate peak inflows to Pond 1 by injecting false data to aggregate the peak outflows of the upstream ponds (Figure 3b). However, this strategy requires the attacker to falsify multiple pond measurements, which reduces the magnitude of the false data that can be used per pond. Additionally, physical constraints and the conduit capacity limit the maximum peak flow that the FDI can create. The second type of FDI impact is to attack the sensor related to Pond 1 to maliciously reduce its outflow, as shown in Figure 3c. Considering that all Err^a would now contribute to a single sensor, the impact tends to be larger.

Both types of FDIs aim to increase the water level in Pond 1, which can lead to a higher flood risk due to reduced tolerance toward control errors from weather forecast uncertainties. An attacker can also combine those two attacking strategies to maximize the impact as shown in Figure 3d. Figure 3d indicates a higher water level peak with the FDI-to-all-sensors scenario than uncontrolled and controlled scenarios. Under the FDI scenario, the water level in Pond 1 exceeds its capacity, resulting in flooding.

2.3. Flood Risks Under Cyber-Physical Attacks Quantification

Knowing the potential FDI impact on the CPS of a smart stormwater system (Section 2.2), we further propose a method to quantify flood risks under FDI to address the stochastic nature rooted in sensor noises and weather forecast uncertainties in this section. Given a smart stormwater system ($\mathcal{N}$), the level of sensor noises ($\boldsymbol{\sigma}_s$), weather forecast uncertainty ($\boldsymbol{\sigma}_w$), an attacking strategy ($\mathcal{A}$), an operation strategy ($\mathcal{O}$) and $t \in \mathcal{T}^a$, flood risks under FDI can be decomposed into three terms and quantified by Equation 20.

$$\underbrace{P_c(F_s)}_{P^s_c} = \underbrace{P(E)}_{P_E} \times \underbrace{P\left(\boldsymbol{z}_t^{\mathrm{T}} \cdot \mathcal{G}^{-1} \cdot \boldsymbol{z}_t \le \varepsilon | \boldsymbol{y}^a\right)}_{P_{FDI}} \times \underbrace{\left[1 - P\left(dy^f_{s,t} \le \left(y^{\max}_s - y^{\text{sim},a}_{s,t}\right) | \boldsymbol{y}^a\right)\right]}_{P^s_W} \tag{20}$$

where $P^s_c = P_c(F_s)$ is the risk of pond s getting flooded within $\mathcal{T}^a$ given the occurrence of FDI. The term F_s is an indicator variable (value "1" if pond s is flooded). Flood risks under FDI are decomposed into three terms. The first term $P_E = P(E)$ is the occurrence probability of a storm E that can be calculated from the return period of

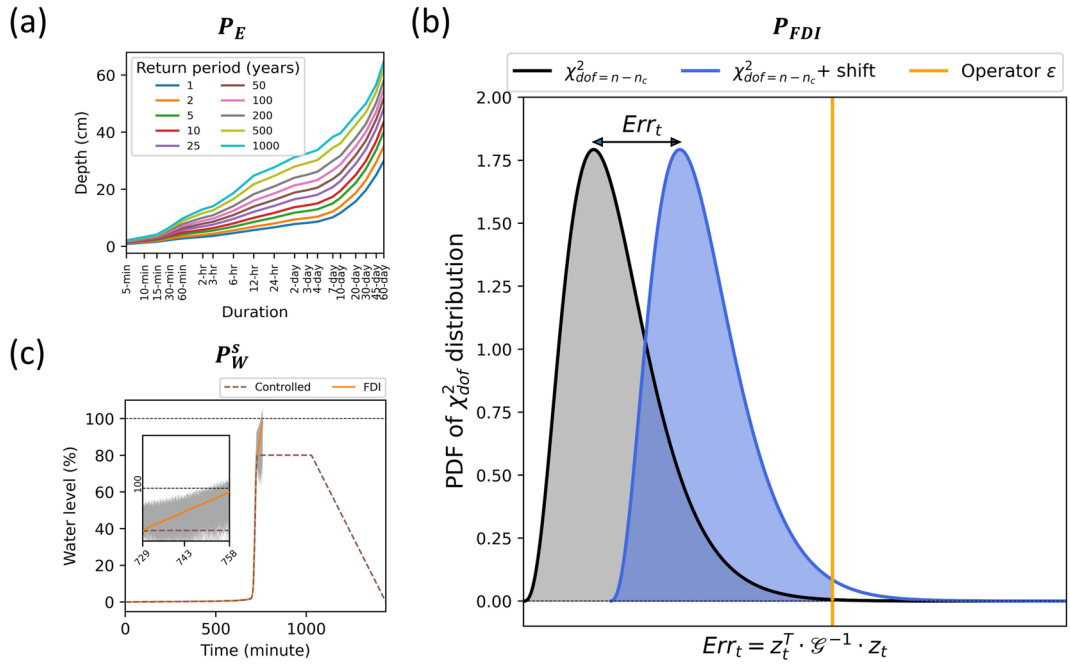

Figure 4. Probability estimation concepts for (a) P_E, (b) P_{FDI}, and (c) P_W^s, where P_E is estimated by DDF curve, P_{FDI} (blue area) is calculated from $\chi^2_{\text{dof}=n-n^a}$ distribution, and P_W^s is computed by Monte Carlo simulations over realizations of sensor noises and forecasted runoffs (gray lines).

a storm. The second term, $P_{\text{FDI}} = P\left(\boldsymbol{z}_t^{\text{T}} \cdot \mathcal{G}^{-1} \cdot \boldsymbol{z}_t \le \varepsilon | \boldsymbol{y}^{\boldsymbol{a}}\right)$, is the successful FDI rate (i.e., FDI occurred but not being detected) given $\boldsymbol{y}^{\boldsymbol{a}}$. The injected false data ($\boldsymbol{y}^{\boldsymbol{a}}$) is solved by optimization introduced in Section 2.2. The third term, $P_W^s = \left[1 - P\left(dy_{s,t}^f \le \left(y_s^{\max} - y_{s,t}^{\text{sim},a}\right) | \boldsymbol{y}^{\boldsymbol{a}}\right)\right]$, is the flooding probability of pond s under the water level control errors ($dy_{s,t}^f$ [cm]) caused by forecast uncertainties given $\boldsymbol{y}^{\boldsymbol{a}}$. The simulated water levels given $\boldsymbol{y}^{\boldsymbol{a}}$ is denoted as $y_{s,t}^{\text{sim},a}$, and $y_s^{\max}$ is the maximum depth of pond s.

To evaluate these three terms, we estimate P_E from frequency analysis using depth-duration-frequency (DDF) curve (Figure 4a). P_{FDI} is calculated by multiplying the blue area under the shifted $\chi^2_{\text{dof}=n-n^a}$ distribution over every time step in $\mathcal{T}^a$ (Figure 4b). The shift is equal to $\text{Err}_t = \boldsymbol{z}_t^{\text{T}} \cdot \mathcal{G}^{-1} \cdot \boldsymbol{z}_t$ as shown in Equation 14. The flooding probability under control errors (P_W^s) is estimated by Monte Carlo simulations, where 1,000 realization sets of sensor noises and forecasted runoffs were generated to evaluate the water level variations given $\boldsymbol{y}^{\boldsymbol{a}}$. For example, even the water level under one single FDI experiment in a deterministic system (orange line in Figure 4c) does not overflow the pond, some realizations of sensor noises and forecast uncertainties might result in floods (some gray lines are over 100% pond's capacity in Figure 4c), where the brown dashed line references the water level in the controlled scenario. Then, P_W^s is computed by the number of flooded realizations to the total number of realizations.

Risks of pond s (P_c^s) getting flooded within T^a, given the occurrence of FDI, is then quantified by multiplying P_E, P_{FDI}, and P_W^s together. Note that the decomposition of P_{FDI} and P_W^s assumes that the runoffs at the previous time step ($\boldsymbol{w}_{t-1}$) is a known value in Equation 5.

3. Materials

3.1. Study Area

This study adopts a stormwater system design (Figure 5) with a network layout of nine ponds, based on a residential area in Bethlehem Township, PA, US. The total basin area is 0.81 km^2. The ponds are radically connected to Pond 1 through circular conduits, with Pond 1 serving as the system outlet that flows into the Nancy Run Creek. The design of the stormwater system was carried out using the US EPA's Storm Water Management Model (SWMM; Rossman, 2010). The goal was to ensure that all pond outflows remained below 1.4 m^3/s during

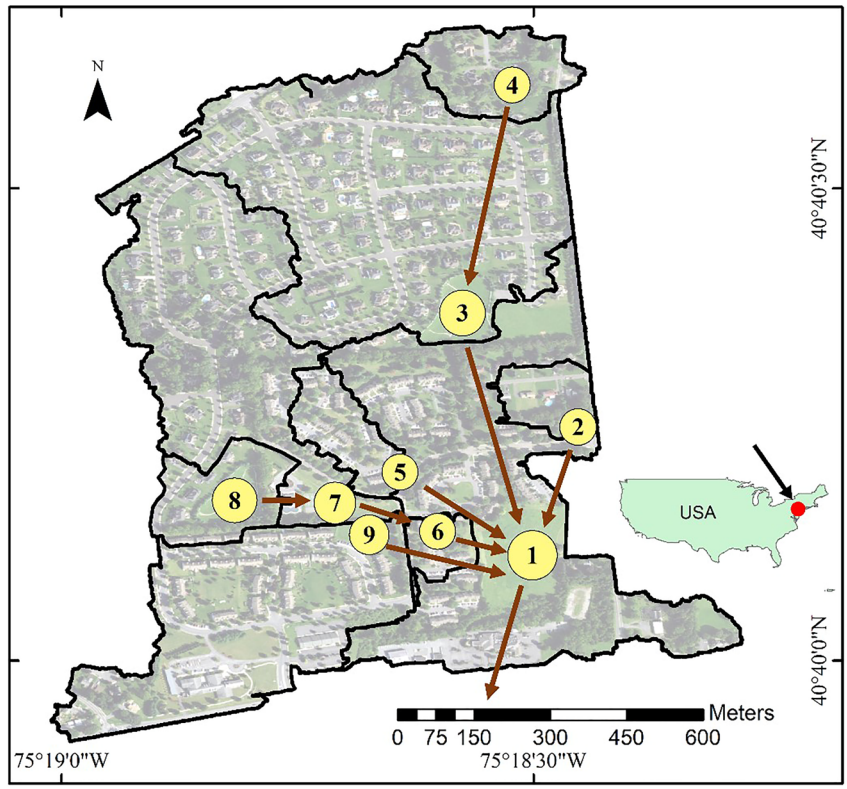

Figure 5. A designed stormwater system with the network layout referenced from a residential area in Bethlehem township, PA, US. Ponds' capacities are indicated by circle size in a log scale.

a 2-year-24-hr design storm (Figure S2 in Supporting Information S1) and that all pond water levels remained below 85% of their storage capacities during a 25-year-24-hr design storm (Figure S3 in Supporting Information S1) in a passive control scenario (i.e., uncontrolled). This was achieved because the designed system was intended to reduce outflow peaks while also ensuring that it could withstand a 25-year-24-hr design storm without causing flooding. The synthetic design storms are based on the standard 24-hr NRCS type II rainfall distribution (NRCS, 2004) with a DDF curve from NOAA Atlas 14 Volume 2 Version 3 (Bonnin et al., 2004). This study only considers 24-hr storms, and all storms mentioned in the content below are 24-hr designed storms. The ponds are assumed to be rectangular for simplicity, with a fixed value for the pond surface area ($\boldsymbol{a}_s$) over depth.

Without losing the generality, the ponds are set to be rectangular, in which the pond's surface area ($\boldsymbol{a}_s$) is a fixed value over depth. A square orifice is located at the bottom of each pond, with the invert elevation of each pond higher than the overflow heights of all downstream ponds to satisfy the gravity-driven outflow assumption (Equation 12). Table S1 in Supporting Information S1 provides detailed information on the configuration of this design stormwater system, including pond capacities, conduit lengths, orifice sizes, and invert elevations. For a "smart" stormwater system, a water level sensor and an outflow gate actuator are installed at each pond, communicating with a SCADA center.

3.2. Numerical Experiment Setup

Our numerical experiment involves three scenarios: (a) uncontrolled, (b) controlled, and (c) FDI associated with the controlled system. We translate the pond-conduit network (Figure 5) into A, B_u, B_w, and C matrixes to establish a state-space model with discretized time step equal to 1 min. First, we calibrate the number of segments (n_c) and gate coefficient ($\boldsymbol{c}_g$) with the SWMM-simulated data in the uncontrolled scenario, where pond outflows are governed by:

$$\boldsymbol{u}_t = \max\left(-\boldsymbol{u}_t^{aw}, -\boldsymbol{u}_t^{uc}\right) \tag{21}$$

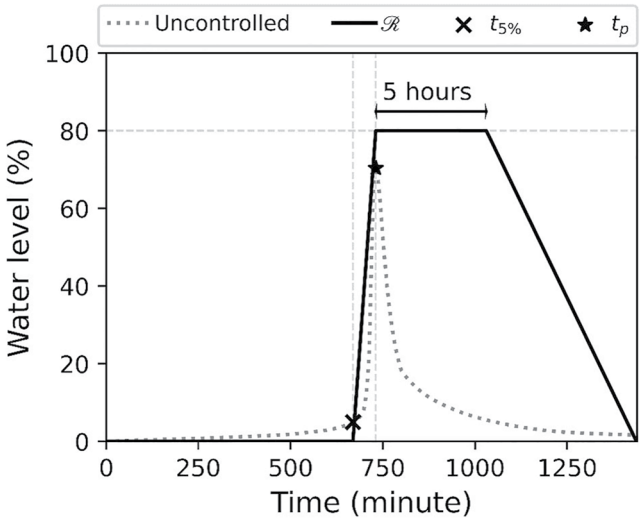

Figure 6. A control rule for desired water level ($\mathscr{r}_t$). $t_{5\%}$ (cross) and t_p (star) are two control time points determined by the simulated water level reaching 5% pond capacity and the peak under an uncontrolled scenario. The desired water level linearly increases from 0% to 80% pond capacity and maintains for 5 hr before decreasing.

For n_c, we search in [1, 4] from upstream to downstream to minimize the averaged Root Mean Square Error (RMSE) of SWMM-simulated water levels under 2-year and 25-year storms (RMSE_{n_c}). Then, we manually tune $\boldsymbol{c_g}$ to minimize the averaged RMSE of SWMM-simulated water levels and pond outflows under 2-year and 25-year storms (RMSE_{C_g}).

Next, we developed a simple control rule for maintaining the desired water level ($\mathscr{r}_t$) in the controlled scenario. The control strategy is designed to retain runoff in ponds by maintaining the water level at 80% of the pond capacity for 5 hr before reducing it. The control rule is illustrated in Figure 6, where the two control time points $t_{5\%}$ and t_p are determined based on the simulated water level in the uncontrolled scenario. Here, $t_{5\%}$ is the time step at which the simulated water level reaches 5% pond capacity, and t_p is the time step when simulated water level reaches its peak. The objective is to gradually accumulate water to 80% of the pond's capacity during the rising period of the water level, which is defined as $t_{5\%}$ to t_p. It is important to note that this control rule is only applied to Ponds 2 to 9. Pond 1, which serves as the system outlet, is designed to drain the water with the maximum physical capacity (as described in Equation 21) for safety reasons, and therefore, the desired water level in this pond is set to zero. To achieve a better control outcome, we tuned the values of R and Q in Equation 8 to satisfy the specific control requirements (refer to Text S2 in Supporting Information S1 for details). However, it is worth noting that finding an optimal control rule is not the focus of this paper.

Last, regarding the FDI scenarios, we consider two cases (a) attacking only the sensor of the targeted pond (FDI_S) and (b) attacking multiple sensors of the targeted pond and its upstream ponds (FDI_M). In Equation 16, we set the values of p and p^a to be 0.95 and 0.99, respectively. We design the attack starting at t_p, in which each pond has a different attacking start time. Also, to prevent solving a large-scale optimization problem and avoid potential numerical issues, we sequentially solved the Mixed-Integer Quadratic Convex Programming (MIQCP) problem (Section 2.2.2) every 5 min. For example, the 30-min-FDI problem was divided into six subproblems.

To estimate P_W^s using Monte Carlo simulations, we generate 1,000 realizations of sensor noises and forecasted runoffs (i.e., $\{\boldsymbol{v}_t, \tilde{\boldsymbol{w}}_t; t \in \mathcal{T}\}$). We sampled sensor noises from a Gaussian distribution with zero mean and standard deviation ($\boldsymbol{\sigma_s}$ [cm]) of 0.25 (MaxBotix MB7384, 2023). Additionally, we synthesize forecasted runoffs using a multiplicative error model that is commonly used in radar forecast literature (Schleiss et al., 2020):

$$\boldsymbol{w}_t = \beta \times \tilde{\boldsymbol{w}}_t \times \boldsymbol{\varepsilon}_w \tag{22}$$

where β is set to 1 and $\boldsymbol{\varepsilon_w}$ is a vector of random numbers sampled from a lognormal distribution with a median of 1 and a standard deviation of 0.9 (Schleiss et al., 2020). Since we assume a time-invariant LQG control system (i.e., K, K_w, and K_r are fixed constant over a storm event), the standard deviation of $\boldsymbol{\varepsilon_w}$ ($\boldsymbol{\sigma_w}$ [cm]) was estimated by averaging the standard deviation calculated at each time step for each pond. We simulate runoffs ($\boldsymbol{w}_t$) using the SWMM model with design storms of various return periods, including 1, 2, 5, 10, 25, 50, 100, and 200 years.

3.3. Control System Properties—Kalman Gain

The significance of Kalman gain (L) in determining the vulnerability of the system to FDI can be understood from Equation 5. This is because L decides the extent to which the false data injection could potentially affect the outflow controls ($\boldsymbol{u}_t$) by modifying the estimates ($\hat{\boldsymbol{x}}_t$) obtained from LQE. Hence, we can examine the Kalman gain to have the initial understanding of the system vulnerability to FDI. In Figure 7, we illustrate the Kalman gains for each pond over various combinations of sensor noises ($\boldsymbol{\sigma_s}$) and forecast uncertainties ($\boldsymbol{\sigma_w}$). The height of each polar bar represents the value of Kalman gain, with the highest value (gray) being one and the lowest being zero. The color gradient indicates the level of sensor noise, while the solid black lines mark the value used in this study. The forecast uncertainties increase in the counterclockwise direction, with larger uncertainties associated with larger storms (i.e., greater return periods), according to Equation 22.

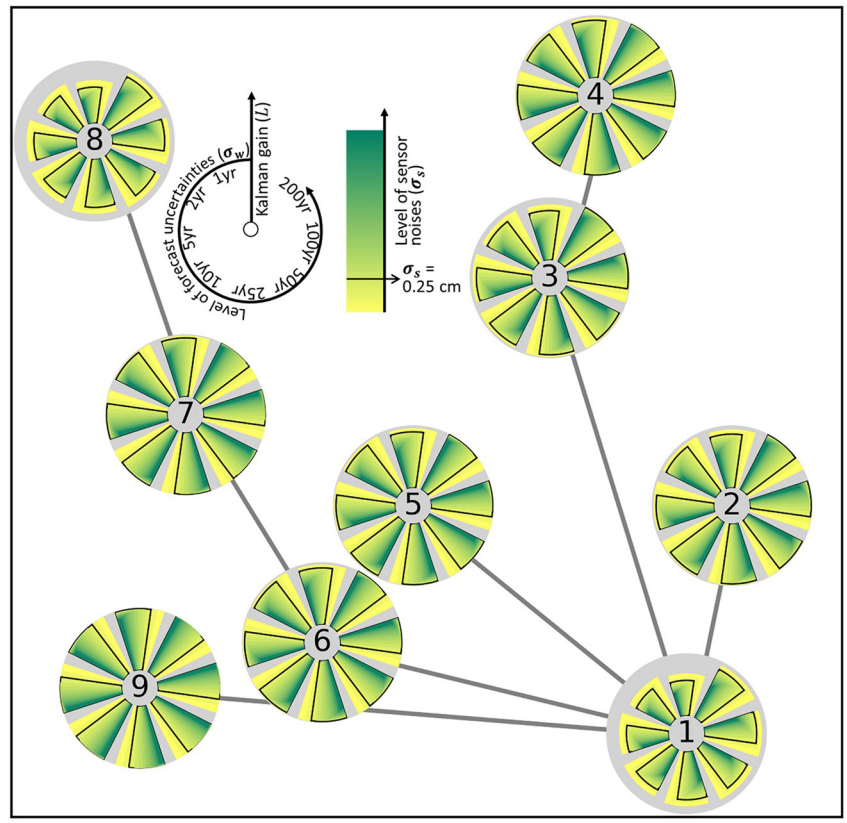

Figure 7. Polar bar chart of Kalman gains (L; heights) over combinations of different levels of sensor noises ($\boldsymbol{\sigma}_s$) and forecast uncertainties ($\boldsymbol{\sigma}_w$) for each pond. The largest L is one (gray). The color gradient indicates the level of $\boldsymbol{\sigma}_s$, where the black solid lines mark the value used in this study. The term $\boldsymbol{\sigma}_w$ increases in the counterclockwise direction.

The results show that most ponds have high L values, indicating a strong dependence on sensor measurements. Ponds 1, 3, and 8 have relatively smaller L values, which suggests that injected false data may have a limited ability to cause floods in these ponds. However, to accurately assess the actual impacts and flood risks, it is necessary to apply the proposed method.

4. Results

4.1. Flood Risks Solely Due To Cyber-Physical Attacks

This section presents the results of the FDI impact analysis conducted prior to quantifying cyber-physical attack risks. The calibrated state-space model has RMSE_{n_c} and RMSE_{C_g} equal to 4.49 and 2.30 cm, respectively. We compare the water level dynamics of the calibrated model under different scenarios, where the water levels under FDI_S and FDI_M are computed by solving MIQCP problems. Figure 8 shows the water level comparisons of uncontrolled (dotted blue line), controlled (dashed brown line), FDI_S (orange line), and FDI_M (green line) with 25-year storm and 30-min attack. Each plot is one independent experiment with different targeted ponds. The nested stem plots indicate relative maximum water level differences between controlled and two FDI scenarios of nine ponds. Lastly, the red lines represent the control target (i.e., the desired water levels).

The control scenario involved detaining the water in the pond for 5 hr to reduce the peak outflow, resulting in a decrease in most ponds' peak outflows (−0.006 to −0.218 m^3/s) except for Pond 8 (+0.001 m^3/s). Pond 1 showed the highest reduction (−0.218 m^3/s) due to water detentions in the upstream ponds. This reduction is also reflected in Pond 1's lower water level in the controlled scenario, as shown in Figure 8. Since Pond 1's desired water level was set to zero (red line), it always had the maximum possible outflow. Therefore, a lower water level corresponded to a lower outflow. These findings are consistent with previous studies on smart stormwater systems (Wong & Kerkez, 2018), which suggest that actively controlling a stormwater system can improve the

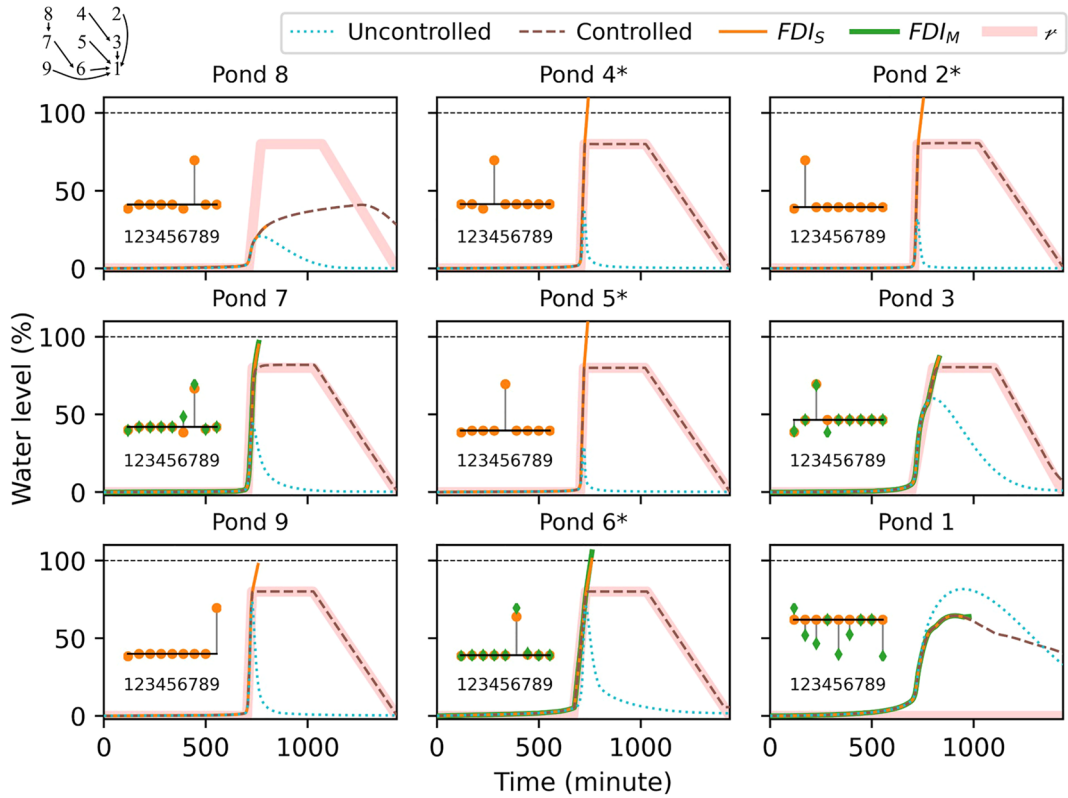

Figure 8. Water level comparisons of uncontrolled (dotted blue line), controlled (dashed brown line), FDI_S (orange line), and FDI_M (green line) with 25-year storm and a 30-min attack. The control targets are shown in red lines. The nested stem plots indicate relative maximum water level differences between controlled and two FDI scenarios of nine ponds. The star signs in the subtitle indicate the flooded ponds under FDI. The pond-conduit network is visualized in the upper left corner.

utilization of existing infrastructure and accommodate larger storms while reducing erosion impact downstream (i.e., lower peak outflow).

However, we demonstrate FDI can lead to floods in a smart stormwater system shown in the results of FDI_S. In Figure 8, Ponds 2, 4, 5, and 6 are flooded (title with * in the figure) if they are attacked, while Ponds 1, 3, and 8 show limited water level rise. These results correspond to the Kalman gain properties revealed in Figure 7. In addition, the physical properties of each pond also contribute to these results. For example, Pond 8 has a relatively large capacity for its inflows, making it difficult to accumulate enough water to reach the desired water level represented by the red line. Pond 1 has the largest capacity, meaning that the same volume change may cause only a small variation in water level compared to the upstream ponds.

We conducted further tests on the FDI_M scenario in a pond that receives inflows from upstream ponds, namely, Ponds 1, 3, 6, and 7. While FDI had a limited impact on reducing Pond 1's outflow, we observed slightly higher water levels in FDI_M than FDI_S. This indicates that larger inflows were generated from upstream ponds due to FDI. This observation is further supported by the lower peak water levels in the stem plot, as water is released downstream. We also noticed a similar trend in Pond 3, where the green dot is lower in Pond 4 (upstream pond). Although the inflows in Ponds 6 and 7 were not significantly increased, the water levels in FDI_M are higher than FDI_S. These results are attributed to the wider attackable range (Err^a) resulting from attacking multiple sensors (i.e., lower degrees of freedom in χ^2 distribution; Equation 16).

In Pond 8's stem plot, we found that the downstream pond, Pond 7, has the largest increased peak water level. This is because contaminated measurements in Pond 8 also affected Pond 7's outflow control. In other words, Pond 7 is mistakenly convinced that no large inflow was coming from the upstream pond, thus reducing its outflow. Although we have explained the reasons behind these control results, it does not necessarily mean that we can address each pond's issue individually. When refining the control system, we need to view it as one

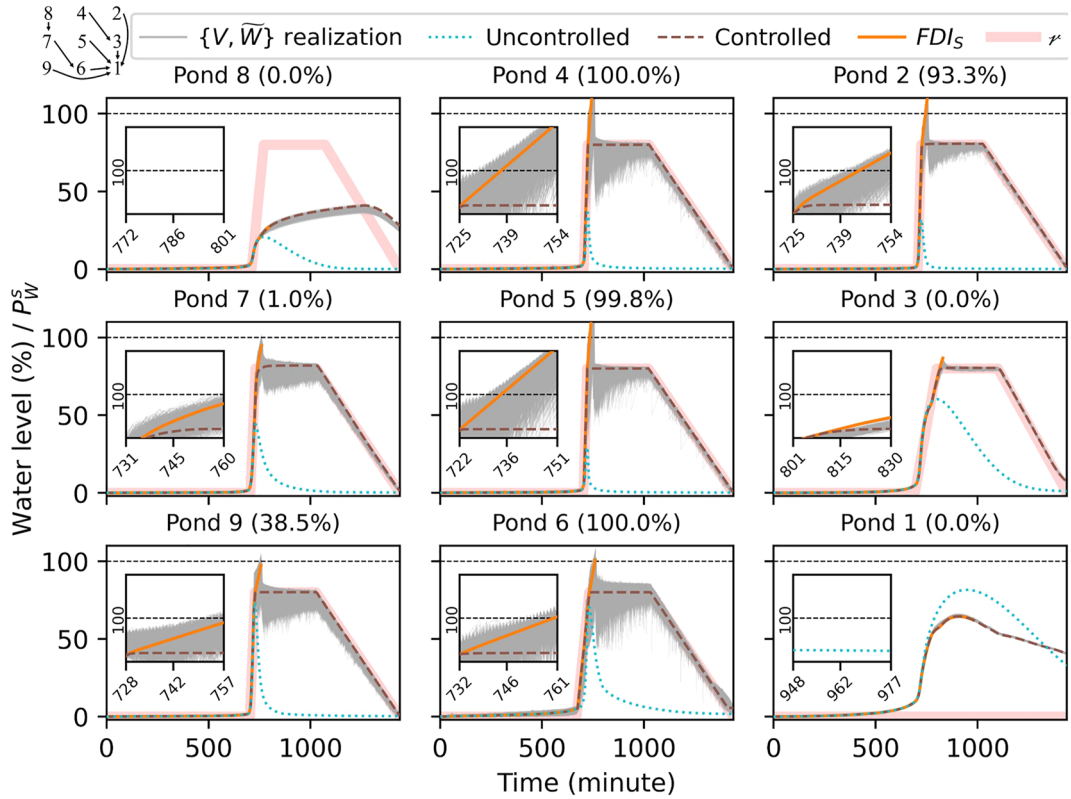

Figure 9. Water level variations under sensor noises and forecasted runoffs with a 25-year storm. Water levels in uncontrolled, controlled, and FDI_S scenarios are shown in dotted blue lines, dashed brown lines, and solid orange lines, respectively. The control targets are shown in red lines. Each plot is one independent experiment differing in attacking targets. The pond-conduit network is visualized in the upper left corner.

networked system, and trade-offs may exist. Given the mild differences between FDI_M and FDI_S in Figure 8, we only consider FDI_S for the following experiments in this case study.

4.2. Flood Risks of Sensor Noises, Forecast Uncertainties, and Cyber-Physical Attacks

In this section, we analyze the impact of sensor noises and forecast uncertainties (i.e., $\{\boldsymbol{v}_t, \tilde{\boldsymbol{w}}_t; t \in \mathcal{T}\}$) on flood risks via Monte Carlo simulations. Figure 9 has a similar layout as Figure 8, but the nested plots are the magnified section for attacking periods, and the percentage of 1,000 realizations (represented by gray lines) that experience flooding is shown in the subplot title.

We observe that Ponds 4 and 6 experience flooding in all realizations, with or without sensor noises and forecast uncertainties. In Ponds 2 and 5, the percentage of floods is not 100% ($P_W^2 = 93.3\%$ and $P_W^5 = 99.8\%$) compared to the deterministic scenario (Figure 8). This could be because some realizations have overestimated runoff forecasts, leading the control system to increase the outflow, which unintentionally neutralizes the effect of FDI. While it may seem like sensor noises and forecast uncertainties reduce flood risks, the probability of over 90% flooding is still high. On the other hand, the effect of sensor noises and forecast uncertainties can increase the flooding probability of Ponds 7 and 9 ($P_W^7 = 1\%$ and $P_W^9 = 38.5\%$) compared to the deterministic scenario. In sum, when considering the effect of sensor noises and forecast uncertainties, the flood risks (i.e., the number of ponds experiencing flooding) increase in our case study.

4.3. Flood Risks of Cyber-Physical Attacks Under Different Storm Return Periods

With the aid of Figure 9, flood risks can be computed using Equation 20. As an example, for Pond 9, the P_E for the 25-year designed storm is 0.04. Since the attacker uses all attackable range at each attacking time step ($\text{Err}_t = \text{Err}^a$ for $t \in T^a$) in this particular case, the successful FDI rate (P_{FDI}) is equal to the designed successful FDI rate (i.e.,

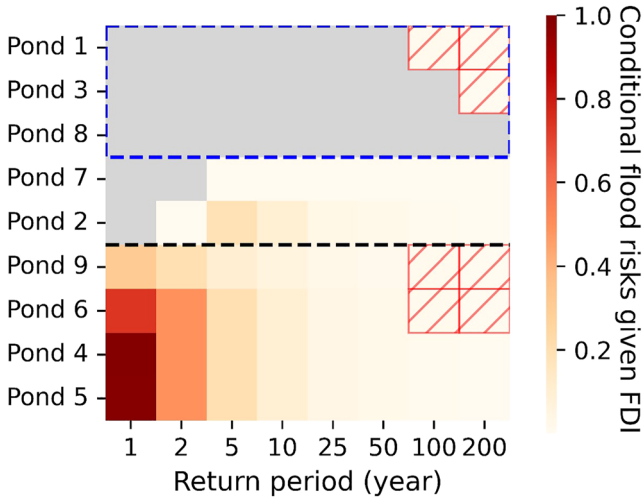

Figure 10. Heatmap of flood risks under cyber-physical attacks over different return periods of 24-hr design storms with attacking lengths (T^a) of 35 min. The red boxes indicate the corresponding flood risks are over 90% contributed by the storms. The gray color indicates zero flood risks. The blue dashed line encloses the ponds that have lower Kalman gain (L).

$p^a = 0.99$). Next, P_W^9 is 0.385 as we shown in the previous section. Finally, flood risks under FDI, P_c^9, can be calculated as $0.04 \times 0.99 \times 0.385 = 0.015$.

Following the quantification procedure, flood risks were determined for various design storm return periods with an attacking length of 35 min ($T^a = 35$), as shown in Figure 10. The darker colors correspond to higher risks, while the gray color represents zero flood risks. The red boxes denote flood risks that are more than 90% contributed by storms. Based on the risk pattern, Figure 10 is divided into three regions. The region enclosed by a blue dashed line indicates that the flood risks are primarily caused by storms, and the pattern is consistent with the intuitive understanding that larger storms result in more significant risks. Ponds 1, 3, and 8 are less affected by FDI due to their lower Kalman gain (see Section 4.2) and physical characteristics.

On the other hand, the other two regions, divided by black dashed lines, have counter-intuitive flood risk patterns. Flood risks are higher with smaller and more frequent storms. This pattern suggests that FDI triggers floods that were not supposed to occur as significant and affects Ponds 2, 4, 5, 6, 7, and 9. Additionally, the flood risks shown in Figure 10 are conditional on the occurrence of FDI. For instance, flood risks are zero with a 1-year return period storm due to the deficiency in stormwater supply in Ponds 2 and 7 (i.e., the second region). However, given the occurrence of FDI, flood risks abruptly increase when surpassing the stormwater supply limits (e.g., 2-year and 5-year return period storms) and gradually decrease with large but less frequent storms. The decreasing patterns can be explained by the lower occurrence probability of larger storms. The third region (i.e., Ponds 4, 5, 6, and 9) is similar to the second region. The difference is that even a 1-year return period storm can cause floods under FDI. As a result, the sudden increase in flood risks was not observed in the third region. Additionally, the impact of different attack lengths (T^a) on flood risks is examined. Flood risks generally increase with a longer attack length (Figure S4 in Supporting Information S1). We also compare the flood risks with and without FDI in Table S2 of the Supporting Information S1.

5. Discussion

5.1. Defending Strategies

This study evaluates flood risks arising from a specific cyber-physical attack termed FDI on process data, particularly the water level. To address this threat, engineers can focus on achieving an optimal balance between control efficiency and error tolerance by modulating the desired water level. Another avenue is to refine the accuracy of the prediction model. Techniques such as the extended Kalman filter (Liu et al., 2016), advancements in weather forecasting, and the adoption of time-variant adjustments to K, K_r, and K_w can be employed to this end. Fundamentally, the efficacy of the FDI hinges on the inherent uncertainty in the control system tied to state estimation.

Nevertheless, these measures do not directly block the intrusion of false data; they merely attenuate its tangible consequences. In parallel, the use of anomaly detection algorithms can differentiate false data from ordinary sensor noise. Several techniques ranging from fast Independent Component Analysis (Brentan et al., 2021), support vector machines (Nader et al., 2016), hidden Markov chains (Zohrevand et al., 2016), to information theory (Ahmed et al., 2016) have been introduced for this task. The realm of data-driven anomaly detection remains a fertile ground for further research (Moazeni & Khazaei, 2022). Yet, it's worth noting that many of these investigations are anchored in contexts like water distribution systems (Taormina et al., 2017). There is a need to characterize sensor data in stormwater systems as their properties vary across storm events. The inherent uncertainties of natural systems can complicate the crafting of an anomaly detection algorithm tailored for stormwater systems.

Moreover, a holistic approach to detecting attacks should encompass not only process data (e.g., water level readings) but also traffic data. Even though this study does not delve into traffic data—the digital pulse that bridges devices, sensors, and controllers—this information can offer insights into irregular patterns that might be indicative of cyber-physical threats. For a comprehensive risk assessment, it's essential to contemplate diverse

attack modalities to reveal all potential vulnerabilities in a cyber-physical system. Prior investigations in water distribution systems have developed platforms that assess a spectrum of attack vectors, both on process and traffic data (Murillo et al., 2023; Taormina et al., 2019). These can set a precedent for subsequent endeavors in stormwater systems.

5.2. Limitations

This study lays the groundwork for understanding flood risks in the context of FDI. We employed numerical tests based on design storms, acknowledging that real-life rainfall events are inherently uncertain and complex. Variabilities in rainfall patterns and intervals between storms introduce uncertainties in initial water levels and peak runoff timings. These uncertainties can significantly alter the flood risks tied to FDI. Addressing these nuances is a promising direction for future research.

Our findings operate under the premise that FDI is in play and that potential attackers possess complete knowledge of the system. In essence, our analysis presents a worst-case scenario for 100% FDI occurrence. It will be extremely difficult, if not impossible, to quantify the occurrence likelihood of when such an attack will happen. However, exploring intrusion pathways (Hahn & Govindarasu, 2011; Lippmann & Ingols, 2005) and applying motivation-based analyses (Ngafeeson, 2010) to discern the attackers' incentives might offer valuable insights.

Furthermore, our risk assessment approach currently limited to χ^2 detector when determining P_{FDI}. Yet, this methodology can be molded to suit other detectors with varying probability distributions, preserving the core idea. The risk assessment principles we've outlined here may be applicable to other interconnected water systems, such as irrigation canals (Conde et al., 2021; Durdu, 2010) and multi-reservoir systems (Georgakakos, 1989; Labadie, 2004; Wasimi & Kitanidis, 1983). These referenced studies have illustrated the ability to represent these systems using state-space models. In both scenarios, the water level can serve as a primary state with sensor reading. Gate openings or water releases will be the "inputs" in the control systems. Furthermore, both systems are subject to some uncertainties that require forecasting. For instance, reservoir inflows carry climate uncertainties, while their releases correlate with the dynamics of downstream demand. Similarly, the irrigation canal system grapples with uncertainties stemming from both inflows and irrigation demands. With this setup in place, our risk quantification approach can be effectively applied.

6. Conclusions

With the growing use of smart technologies in water systems, cybersecurity has become a critical concern. Yet, the impact and the risk of FDI attacks in these systems remain understudied. This study proposes a mathematical framework to address this gap, applying it to a nine-pond smart stormwater system with water level sensors and outflow gate actuators. Using a state-space model, we simulate stormwater dynamics, while a real-time LQG controller manages pond outflows. Our methodology assesses the cyber-physical impacts of storms and FDI, quantifying flood risks from both FDI and varying storm intensities.

Our results demonstrate the potential consequences of FDI, such as generating a massive inflow peak or maliciously lowering the outflow to increase flood risks. We further reveal how FDI can drastically alter flood risk patterns across different storm intensities. Understanding these flood risks in the face of FDI can inform system planning and investment decisions. The proposed mathematical framework offers a foundation for evaluating defense strategies against potential attacking strategies in future studies. Despite certain limitations, our quantification method can extend to other networked water systems, including irrigation canal and multi-reservoir systems.

Data Availability Statement

The resources utilized for this research, including the SWMM model, DDF curve, simulation models, and Gurobi optimization models as discussed in Section 2.2.2, are accessible in the Lin et al. (2023) archive. Additionally, the Python script used to produce the figures presented in this document can be found in the same Lin et al. (2023) reference.

Acknowledgments
The work described in this paper was supported by the US National Science Foundation (NSF): CBET 1941727 and a grant from the Commonwealth of Pennsylvania, Department of Community and Economic Development, through the Pennsylvania Infrastructure Technology Alliance (PITA). We would like to thank the editor, the associate editor, and anonymous reviewers for their comments and suggestions to improve the quality of the manuscript.

References

Ahmed, C. M., Murguia, C., & Ruths, J. (2017). Model-based attack detection scheme for smart water distribution networks. In *In proceedings of the 2017 ACM on Asia conference on computer and communications security* (pp. 101–113).
Ahmed, M., Naser Mahmood, A., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, *60*, 19–31. https://doi.org/10.1016/j.jnca.2015.11.016
Amin, S., Litrico, X., Sastry, S., & Bayen, A. M. (2012). Cyber security of water SCADA systems—Part I: Analysis and experimentation of stealthy deception attacks. *IEEE Transactions on Control Systems Technology*, *21*(5), 1963–1970. https://doi.org/10.1109/tcst.2012.2211873
Amin, S., Litrico, X., Sastry, S. S., & Bayen, A. M. (2013). Cyber security of water SCADA systems—Part II: Attack detection using enhanced hydrodynamic models. *IEEE Transactions on Control Systems Technology*, *21*(5), 1679–1693. https://doi.org/10.1109/tcst.2012.2211874
Bartos, M., Wong, B., & Kerkez, B. (2018). Open storm: A complete framework for sensing and control of urban watersheds. *Environmental Sciences: Water Research & Technology*, *4*(3), 346–358. https://doi.org/10.1039/c7ew00374a
Bergal, J. (2021). Florida hack exposes danger to water systems. Retrieved from https://pew.org/3btxWBc
Bonnin, G. M., Martin, D., Lin, B., Parzybok, T., Yekta, M., & Riley, D. (2004). Precipitation-frequency Atlas of the United States. Version 3.0. (Vol. *2*).
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*(6), 795. https://doi.org/10.3390/w13060795
Burian, S. J., & Edwards, F. G. (2002). *Historical perspectives of urban drainage* (pp. 1–16). Global Solutions for Urban Drainage.
Conde, G., Quijano, N., & Ocampo-Martinez, C. (2021). Modeling and control in open-channel irrigation systems: A review. *Annual Reviews in Control*, *51*, 153–171. https://doi.org/10.1016/j.arcontrol.2021.01.003
Depoy, J., Phelan, J., Sholander, P., Smith, B., Varnado, G. B., & Wyss, G. (2005). Risk assessment for physical and cyber attacks on critical infrastructures. In *MILCOM 2005-2005 IEEE military communications conference* (pp. 1961–1969). IEEE.
DePoy, J., Phelan, J., Sholander, P., Smith, B. J., Varnado, G. B., Wyss, G. D., et al. (2006). Critical infrastructure systems of systems assessment methodology. Sandia report.
Durdu, Ö. F. (2010). Fuzzy logic adaptive Kalman filtering in the control of irrigation canals. *International Journal for Numerical Methods in Fluids*, *64*(2), 187–208. https://doi.org/10.1002/fld.2151
Fletcher, T. D., Shuster, W., Hunt, W. F., Ashley, R., Butler, D., Arthur, S., et al. (2015). SUDS, LID, BMPs, WSUD and more–The evolution and application of terminology surrounding urban drainage. *Urban Water Journal*, *12*(7), 525–542. https://doi.org/10.1080/1573062x.2014.916314
Froise, S., & Burges, S. J. (1978). Least-cost design of urban-drainage networks. *Journal of the Water Resources Planning and Management Division*, *104*(1), 75–92. https://doi.org/10.1061/jwrddc.0000086
Gaborit, E., Muschalla, D., Vallet, B., Vanrolleghem, P. A., & Anctil, F. (2013). Improving the performance of stormwater detention basins by real-time control using rainfall forecasts. *Urban Water Journal*, *10*(4), 230–246. https://doi.org/10.1080/1573062x.2012.726229
Georgakakos, A. P. (1989). Extended linear quadratic Gaussian control: Further extensions. *Water Resources Research*, *25*(2), 191–201. https://doi.org/10.1029/wr025i002p00191
Hahn, A., & Govindarasu, M. (2011). Cyber attack exposure evaluation framework for the smart grid. *IEEE Transactions on Smart Grid*, *2*(4), 835–843. https://doi.org/10.1109/tsg.2011.2163829
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*(5), 03120003. https://doi.org/10.1061/(asce)ee.1943-7870.0001686
Huang, Y., Tian, Z., Ke, Q., Liu, J., Irannezhad, M., Fan, D., et al. (2020). Nature-based solutions for urban pluvial flood risk management. *Wiley Interdisciplinary Reviews: Water*, *7*(3), e1421. https://doi.org/10.1002/wat2.1421
Jongman, B. (2018). Effective adaptation to rising flood risk. *Nature Communications*, *9*(1), 1–3. https://doi.org/10.1038/s41467-018-04396-1
Kriaa, S., Pietre-Cambacedes, L., Bouissou, M., & Halgand, Y. (2015). A survey of approaches combining safety and security for industrial control systems. *Reliability Engineering & System Safety*, *139*, 156–178. https://doi.org/10.1016/j.ress.2015.02.008
Kučera, V. (1973). A review of the matrix Riccati equation. *Kybernetika*, *9*(1), 42–61.
Kumar, A., Saxena, N., Jung, S., & Choi, B. J. (2021). Improving detection of false data injection attacks using machine learning with feature selection and oversampling. *Energies*, *15*(1), 212. https://doi.org/10.3390/en15010212
Labadie, J. W. (2004). Optimal operation of multireservoir systems: State-of-the-art review. *Journal of Water Resources Planning and Management*, *130*(2), 93–111. https://doi.org/10.1061/(asce)0733-9496(2004)130:2(93)
Lin, C.-Y., Yang, Y.-C. E., & Moazeni, F. (2023). philip928lin/flood risks of cyber physical attacks in a smart storm water system: Flood risks of cyber-physical attacks in a smart storm water system (v1.0.0) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.10035726
Lippmann, R. P., & Ingols, K. W. (2005). An annotated review of past papers on attack graphs.
Liu, S., Wei, G., Song, Y., & Liu, Y. (2016). Extended Kalman filtering for stochastic nonlinear systems with randomly occurring cyber attacks. *Neurocomputing*, *207*, 708–716. https://doi.org/10.1016/j.neucom.2016.05.060
MaxBotix MB7384. (2023). Retrieved from https://maxbotix.com/products/mb7384?_pos=1&_sid=d81d19b44&_ss=r
Ministry of the Environment. (2003). Stormwater management planning and design manual. *Water Resources*.
Mo, Y., & Sinopoli, B. (2010). False data injection attacks in control systems. In *Preprints of the 1st workshop on secure control systems* (Vol. 1).
Moazeni, F., & Khazaei, J. (2021). Sequential false data injection cyberattacks in water distribution systems targeting storage tanks; a bi-level optimization model. *Sustainable Cities and Society*, *70*, 102895. https://doi.org/10.1016/j.scs.2021.102895
Moazeni, F., & Khazaei, J. (2022). Detection of random false data injection cyberattacks in smart water systems using optimized deep neural networks. *Energies*, *15*(13), 4832. https://doi.org/10.3390/en15134832
Mullapudi, A., Wong, B. P., & Kerkez, B. (2017). Emerging investigators series: Building a theory for smart stormwater systems. *Environmental Sciences: Water Research & Technology*, *3*(1), 66–77. https://doi.org/10.1039/c6ew00211k
Murillo, A., Taormina, R., Tippenhauer, N. O., & Galelli, S. (2023). High-fidelity cyber and physical simulation of water distribution systems. II: Enabling cyber-physical attack localization. *Journal of Water Resources Planning and Management*, *149*(5), 04023010. https://doi.org/10.1061/jwrmd5.wreng-5854
Nader, P., Honeine, P., & Beauseroy, P. (2016). Detection of cyberattacks in a water distribution system using machine learning techniques. In *2016 sixth international conference on digital information processing and communications (ICDIPC)* (pp. 25–30). https://doi.org/10.1109/ICDIPC.2016.7470786
Natural Resources Conservation Service (NRCS). (2004). *National engineering handbook, Part 630-Hydrology*. U.S. Department of Agriculture.
Ngafeeson, M. (2010). *Cybercrime classification: A motivational model*. College of Business Administration, The University of Texas-Pan American. 1201.

Piro, P., Turco, M., Palermo, S. A., Principato, F., & Brunetti, G. (2019). A comprehensive approach to stormwater management problems in the next generation drainage networks. *The Internet of Things for Smart Urban Ecosystems*, 275–304.

Rossman, L. A. (2010). *Storm water management model user's manual, version 5.0* (p. 276). National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency.

Sadler, J. M., Goodall, J. L., Behl, M., Bowes, B. D., & Morsy, M. M. (2020). Exploring real-time control of stormwater systems for mitigating flood risk due to sea level rise. *Journal of Hydrology*, *583*, 124571. https://doi.org/10.1016/j.jhydrol.2020.124571

Schleiss, M., Olsson, J., Berg, P., Niemi, T., Kokkonen, T., Thorndahl, S., et al. (2020). The accuracy of weather radar in heavy rain: A comparative study for Denmark, the Netherlands, Finland and Sweden. *Hydrology and Earth System Sciences*, *24*(6), 3157–3188. https://doi.org/10.5194/hess-24-3157-2020

Schuurmans, J. (1997). Control of water levels in open-channels.

Shin, S., Lee, S., Burian, S. J., Judi, D. R., & McPherson, T. (2020). Evaluating resilience of water distribution networks to operational failures from cyber-physical attacks. *Journal of Environmental Engineering*, *146*(3), 04020003. https://doi.org/10.1061/(asce)ee.1943-7870.0001665

Shishegar, S., Duchesne, S., & Pelletier, G. (2018). Optimization methods applied to stormwater management problems: A review. *Urban Water Journal*, *15*(3), 276–286. https://doi.org/10.1080/1573062x.2018.1439976

Shishegar, S., Duchesne, S., Pelletier, G., & Ghorbani, R. (2021). A smart predictive framework for system-level stormwater management optimization. *Journal of Environmental Management*, *278*, 111505. https://doi.org/10.1016/j.jenvman.2020.111505

Taormina, R., Galelli, S., Douglas, H. C., Tippenhauer, N. O., 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. https://doi.org/10.1016/j.envsoft.2018.11.008

Taormina, R., Galelli, S., Tippenhauer, N. O., Salomons, E., & Ostfeld, A. (2017). Characterizing cyber-physical attacks on water distribution systems. *Journal of Water Resources Planning and Management*, *143*(5), 04017009. https://doi.org/10.1061/(asce)wr.1943-5452.0000749

Taormina, R., Galelli, S., Tippenhauer, N. O., Salomons, E., Ostfeld, A., Eliades, D. G., et al. (2018). Battle of the attack detection algorithms: Disclosing cyber attacks on water distribution networks. *Journal of Water Resources Planning and Management*, *144*(8), 04018048. https://doi.org/10.1061/(asce)wr.1943-5452.0000969

The White House. (2022). Fact sheet: Biden-Harris administration expands public-private cybersecurity partnership to water sector. Retrieved from https://www.whitehouse.gov/briefing-room/statements-releases/2022/01/27/fact-sheet-biden-harris-administration-expands-public-private-cybersecurity-partnership-to-water-sector

Tuptuk, N., Hazell, P., Watson, J., & Hailes, S. (2021). A systematic review of the state of cyber-security in water systems. *Water*, *13*(1), 81. https://doi.org/10.3390/w13010081

Van Meter, R. J., Swan, C. M., & Snodgrass, J. W. (2011). Salinization alters ecosystem structure in urban stormwater detention ponds. *Urban Ecosystems*, *14*(4), 723–736. https://doi.org/10.1007/s11252-011-0180-9

Wasimi, S. A., & Kitanidis, P. K. (1983). Real-time forecasting and daily operation of a multireservoir system during floods by linear quadratic Gaussian control. *Water Resources Research*, *19*(6), 1511–1522. https://doi.org/10.1029/wr019i006p01511

Wong, B. P., & Kerkez, B. (2018). Real-time control of urban headwater catchments through linear feedback: Performance, analysis, and site selection. *Water Resources Research*, *54*(10), 7309–7330. https://doi.org/10.1029/2018wr022657

Yeh, C. H., & Labadie, J. W. (1997). Multiobjective watershed-level planning of storm water detention systems. *Journal of Water Resources Planning and Management*, *123*(6), 336–343. https://doi.org/10.1061/(asce)0733-9496(1997)123:6(336)

Zohrevand, Z., Glasser, U., Shahir, H. Y., Tayebi, M. A., & Costanzo, R. (2016). Hidden Markov based anomaly detection for water supply systems. In *2016 IEEE international conference on big data (big data)* (pp. 1551–1560). IEEE.