WTC^2: Impact-Aware Threat Analysis for Water Treatment Centers

Amarjit Datta
Tennessee Technological University
Cookeville, USA
Email: datta.amarjit@gmail.com

Mohammad Ashiqur Rahman
Florida International University
Miami, USA
Email: marahman@fiu.edu

Hossain Shahriar
Kennesaw State University
Kennesaw, USA
Email: hshahria@kennesaw.edu

***Abstract*—A water treatment center (WTC) removes contaminants and unwanted components from the water and makes the water more acceptable to the end-users. A modern WTC is equipped with different water sensors and uses a combination of wired/wireless communication network. During the water treatment process, controllers periodically collect sensor measurements and make important operational decisions. Since accuracy is vital, a WTC also uses different data validation mechanisms to validate the incoming sensor measurements. However, like any other cyber-physical system, water treatment facilities are prone to cyberattacks and an intelligent adversary can alter the sensors measurements stealthily, and corrupt the water treatment process. In this work, we propose WTC Checker (WTC^2), an impact-aware formal analysis framework that demonstrates the impact of stealthy false data injection attacks on the water treatment sensors. Through our work, we demonstrate that if an adversary has sufficient access to sensor measurements and can evade the data validation process, he/she can compromise the sensors measurements, change the water disinfectant contact time, and inflict damage to the clean water production process. We model this attack as a constraint satisfaction problem (CSP) and encode it using Satisfiability Modulo Theories (SMT). We evaluate the proposed framework for its threat analysis capability as well as its scalability by executing experiments on different synthetic test cases.**

***Index Terms*—Water Treatment Center; Security; Formal Methods; Threat Analysis**

I. Introduction

Water treatment is the process of removing undesirable elements from the water and making water safe for its end-users. Raw water contains harmful chemicals, biological contaminants, and suspended solids. Purified water can be used for drinking, medical purposes and pharmacological, chemical, and industrial applications. Based on the final application of the clean water, the raw water goes through multiple purification processes. In the United States, a combination of ultra-filtration techniques, pH adjustment, ozonation, chlorination, and reverse osmosis (RO) techniques are applied to produce clean drinkable water [1][2].

During the water treatment process, raw water is mixed with different disinfection agents. The amount of disinfects required for water purification depends on multiple water factors such as temperature and pH level. For the optimal operation of the water treatment center (WTC), calculating the required amounts of disinfects is very important. Modern WTCs are equipped with sensors that measure critical properties such as pH level, water temperature, and chlorine level. A WTC is a sophisticated cyber-physical system with physical elements (treatment channels, water tanks) and cyber elements (sensors, controllers, and other internal communication interfaces) [3].

In this work, we propose a impact-aware formal framework that analyzes the impact of cyberattacks on the water treatment sensor measurements. More specifically, the framework analyzes the possibility of unidentified false data injection (UFDI) attacks on the water treatment pH and temperature sensors and measures their impact in the clean water production process. Our prime contributions are as follows:

- We have proposed WTC Checker (WTC^2), a formal attack analysis framework that measures the impact of UFDI attack on the sensors measurements of the water treatment process. When measuring the impact of the attack, we focus on two important benchmarks: (1) treatment time and (2) volume.
- We have formally modeled the water treatment process as a multi channel, time-driven, assembly line where as each step certain water purification subtask take place.
- When modeling the UFDI attack, we have considered important attack properties such as adversary's resource, sensor security, and adversary's accessibility. In this work, we have also formalized a threshold driven data validation process.
- We encode our formalization in SMT [4], which is a powerful theorem solver. We provide a case study to illustrate the execution of the framework. We also evaluate the framework in terms of its threat analysis capability and its scalability.

The rest of this paper is organized as follows: In Section II, we provide necessary background, our motivation and contributions, and related literature works.The proposed framework is briefly discussed in Section III. Formalization of the attack model is presented in Section IV. In Section V, we present two illustrative examples of the execution of our framework. Evaluation results are discussed in Section VI. We conclude the paper in Section VII.

II. Background, Objective, and Related Works

In this section, we provide necessary background related to water treatment process. Then we discuss the motivation (potential attacks on sensors) and our contributions. We also briefly discussed the related work.

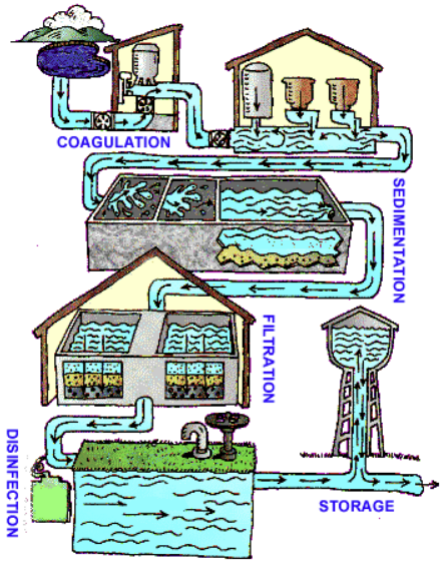


Fig. 1. Raw water treatment process [5].

A. Water Treatment Process

Raw water goes through rigorous purification processes based on the initial condition of the water source and the requirements of the end-users. Substances that are removed from the raw water include parasites (such as Giardia or Cryptosporidium), bacteria, algae, viruses, fungi, minerals (including toxic metals such as lead, copper etc.), and man-made chemical pollutants. We can observe the raw water treatment process in Fig 1. In the following, we discuss some of the common water treatment procedures:

- *Coagulation and Flocculation*: In coagulation and flocculation, chemicals with a positive charge are added to the water. The positive charge neutralizes the negative charge of the dirt and other particles and forms larger particles, called floc.
- *Sedimentation*: During this process, flocs generated from the coagulation and flocculation steps settles in the bottom of the water tank due to their relative weight.
- *Filtration*: After sedimentation, clear water on the top passes through a series of filters composed of sand, gravel, and charcoal. Filtration removes dust, parasites, bacteria, viruses, and chemicals from the water.
- *Disinfection*: Water is disinfected using either the chlorination or the ozonation process. The disinfection step is critical to remove the remaining parasites, bacteria, and viruses from the water and make the water safe for drinking. For chlorine in the Chlorination process, compressed elemental gas, sodium hypochlorite solution (NaOCl) or solid calcium hypochlorite is used [6].

Water pH Requirement: pH is a measure of the hydrogen ion concentration of a solution. pH value ranges from 0 to 14. A pH value of 7 is neutral. If the pH value of the water is less than 7, the water is acidic. Acidic water is bad for health and also hinders the water treatment and damages distribution pipes. If the pH value of the water is more than 7, the water is alkaline. Alkaline water tastes bad. For drinking water, the normal pH range is 6 -to 8.5 [7]. pH value of the water effects the water disinfection process. If the pH value of the water is low, it requires less volume of chlorine to disinfect and a smaller contact time. On the other hand, low pH water is harmful for human consumption. For this reason, the water pH level must be adjusted accordingly before and after the water treatment process. The first pH adjustment happens before the disinfection process (to control the amount of required disinfect) and the second pH adjustment happens after the disinfection process (to make the water more acceptable for human consumption).

Water Disinfection Level and CT Value: The primary disinfectansts for water are chlorine, ozone, and UV light. A main focus of the water disinfection process is the removal of the parasites such as Giardia from the raw water. According to the United States Environmental Protection Agency (EPA), there cannot be more than one Giardia cyst in 100,000 liters of treated water [8]. The WTC's target is the removal of Giardia cysts up to the target limit. If the raw water contains 10,000 Giardia cysts in 100,000 liters of water then the water disinfection process must reduce the concentration of the Giardia by 10,000 times to meet the EPA target. Another way of expressing this requirement is by using the Log reduction level. Since 10,000 has four zero-'es, the water disinfection process must disinfect the water four Log reduction level to meet the disinfection target. In the United States, the minimum Giardia inactivation requirement is 3.0 Logs. We use parameter L to represent the Log removal value.

In the chlorination process, free chlorine is used to disinfect the microorganisms present in the water. Effectiveness of disinfection depends on the microorganism, the concentration of the disinfectant, the contact time, and the temperature and pH value of the water. The term CT is used to denote the product of the concentration of chlorine present in the water and the time the water is in contact with that chlorine. CT value is typically expressed in units of mg-min/L. In the water treatment process, it is important to determine the appropriate value of CT.

Sensor Measurements and CT Value Calculation: To calculate the CT value, the following measurements are needed:

- Temperature: In the chlorination process, as the temperature increases, more chlorine concentration is needed for longer contact time to disinfect the water.
- pH valuer: As the pH value of the water increases, chlorine's ability to disinfect water decreases. Accurate measurement of the pH level is needed to determine the required chlorine concentration and contact time.

The equation for calculating the CT value is as follows [9]:

$$CT = .2828 \times (\bar{ph}^{2.69}) \times (cl^{.15}) \times (.933^{(t-5)}) \times L \quad (1)$$

Here, parameter pH is the pH of the water, cl is the free chlorine residual, mg/l, t is the temperature in degrees, and L is the Log removal.

B. Cyberattack on Sensors

In a cyber-physical system, controllers use sensor measurements to operate its activities optimally. If the sensors are compromised, controllers will not be able to measure the system properties accurately. An adversary can compromise a sensor in various ways. If an adversary has physical or remote access to the sensor terminal, he can corrupt the sensor. Some sensors do not have strong embedded security installed. Since most of the sensors operate in rough terrain, they are expected to have longer battery life, hence have a very low computational ability. As a result, an adversary can easily intercept sensor transmissions and perform cyberattacks such as man-in-the-middle or denial-of-service. Wireless sensors suffer from the jamming attack where an adversary can create artificial contention on its transmission channel. If an adversary can capture a sensor packet, it can perform other cyberattacks such a type-flaw attack, reply attack, or denial-of-service. In this work, we have not discussed any actual attack process. Rather, we focused on the impact of an attack (if the attack really takes place).

C. Our Contributions

We can observe from Equation 1 that the CT value calculation depends on the correct measurement of the pH level and the temperature of the water. A WTC may use different techniques to validate its sensor measurements. However, if an adversary can evade the data validation process, he/she can launch a UFDI attack on the pH and temperature sensors and corrupt the CT calculation. Therefore, there is a great need to explore the possibility of a UFDI attack on the water treatment sensors. We present WTC^2, a formal framework for analyzing the impacts of UFDI cyberattack on the water treatment pH and temperature sensors. In our work, we have used a comprehensive model of the water treatment process, realistic attack attributes, and adversary's capabilities.

An adversary wants to corrupt the sensor measurements and interrupt the purification process. By corrupting the pH and temperature sensor measurements, an adversary wants to jeopardize the CT calculation and affect the disinfection process. Such cyberattacks have two major impacts. The first impact is the reduction of clean water production. The second impact is the delay induced due to imperfect purification of raw water. Using our proposed framework, we can identify all possible attack scenarios. Our framework can return, (1) all attack vectors, (2) best attack vector (in terms of inflicted damage), (3) critical sensors, and (4) overall performance of the system in presence of cyberattacks.

D. Related Works

The concept of a stealthy attack on the power-grid was first presented by Liu et al. in [10]. In [11], the concept of a UFDI attack was further extended by considering limited access to power-grid meters and,-limited attack resources, and assumed that the adversary has complete information about the grid. Ericsson et al. discussed the vulnerability of the SCADA system, when connected with the network for the first time in [12]. In [13] [14], the authors addressed the existing vulnerabilities with the SCADA in the power grid. Haimes et al. in [15] discuss the possibility of cyberattacks on the public water system and proposed a hierarchical holographic model that models multiple perspectives on the hardening of water systems. In [16], Ezell et al. presented a infrastructure risk analysis application that can model the water distribution system of a small municipality and can simulate the water contamination attack scenario. In [17], Gao et al. proposed a formal vulnerability analysis framework for AC state estimation of power grid. In their approach, the authors also use SMT to encode their power grid model. Although, the authors applied similar techniques, they did not consider important attack properties such as stealthiness of the cyberattack, sensor security, total attack resource, and adversary's accessibility, which made our work completely different than theirs.

Among all cyberattacks on the water system, the Maroochy Water Services breach attack is the most well researched and widely investigated. In [18][19], the authors presented the Maroochy Water Breach case. In their work, the authors explained how the cyberattack on the SCADA system in the Maroochy water breach happened. In March 2000, Maroochy wastewater system experienced a cyberattack in which the communications sent by radio links to wastewater pumping stations were being lost, pumps were not working properly, and alarms put in place to alert staff to faults were not going off. During the three-month attack period, the adversary released one million liters of untreated sewage into a stormwater drain from where it flowed to local waterways.

In [3], Adepu et al. investigated the impact of single-point cyberattacks on water system testbed named Secure Water Treatment (SWaT). In their setup, SCADA system is connected with the Programmable Logic Controller (PLCs) servers, that in turn are connected with the sensors and actuators of the water treatment system. With single-point-attacks, the authors presented how the attack propagates between sensors and actuators. Panguluri et al. in [20] presented some statistics on cyberattacks and resulting damages on the water and waste water treatment infrastructure. The authors discussed why maintaining cybersecurity on the critical infrastructure is difficult and explained what measures authorities can take to counter cyber-threats. In [21], Kang et al. proposed a formal technique based on constraint solving for analyzing cyberattacks on water treatment plants. In their work, an adversary can compromise a WTC's sensor measurements by compromising the lists relayed from the PLC to the sensor and from the PLC to the actuator. However, in their work, the authors did not consider multi channel attacks where each channel may contain an independent set of sensors. Also in this work, the authors did not consider sensor security, accessibility, and the adversary's resources.

Cyberattack on WTCs is an important research domain. Although some of the above literature works are compelling, none of them discusses the impact of cyberattacks on the sensors of the water treatment system. In this work, we have analyzed the impact of cyberattacks on the water treatment

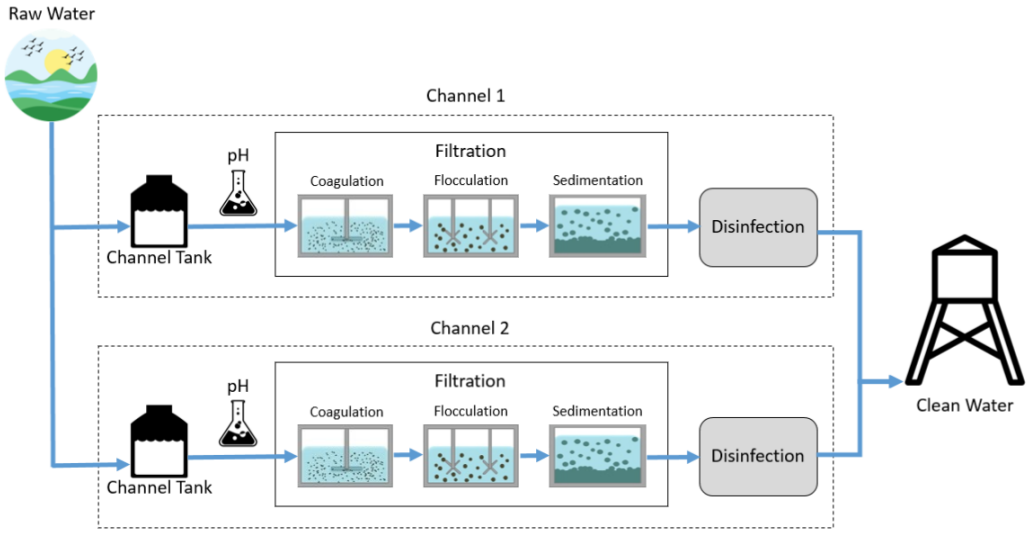


Fig. 2. An example water treatment center.

sensors and measures their implication on the CT value calculation. We also quantify the impact of the attack in terms of the water treatment time and the clean water volume, which is unique to the best of our knowledge. While it appears intuitive that an attack on the water treatment sensors can compromise the overall clean water production quality of the WTC, we provide a systematic modeling framework to analyze such cyberattacks.

III. Attack Model

We perform threat analysis considering a stealthy attack model as discussed below.

A. Attack Attributes

An adversary wants to launch cyberattack on water treatment sensors. To launch cyberattacks, an adversary must have access to the sensors. In a WTC, there can be more than one treatment channels. Multiple channels are important for the seamless operation of the water treatment facility. If one channel undergoes maintenance, the other channel can serve the end-users and produce clean water for the end-users. From Fig 2, we can observe an architecture of a typical WTC with three channels. Each channel operates independently. Multiple channels are necessary to ensure the reliable operation of the WTC. pH and temperature sensors measurements of different channels may produce similar but different values based on the cleanliness of the channel.

We define the UFDI attack on the water treatment process in terms of its attributes as follows:

- The target of an adversary is to attack the water treatment sensors, deceive the data validation mechanism, and reduce the clean drinking water production. An adversary also wants to introduce delay in the water treatment process and make the entire process inefficient.
- An adversary can only attack a sensor if the sensor is not secured and accessible. In this work, we use the term accessible to denote the adversary's capability to physically or remotely access the control terminal of the sensor. Similarly, the term secured denotes whether the sensor measurements are encrypted and an adversary has enough capability to perform a false data injection attack.
- An adversary cannot attack all the channels simultaneously. At a particular time, an adversary can attack limited (k) number of channels.
- An adversary cannot attack the same channel continuously. To remain stealthy, an adversary must wait at least y time slots before it can attack the same channel again. We denote the parameter y as consecutive attack threshold.
- During the entire attack, an adversary can intrude a maximum z number of times. Here, z denotes the adversary's resource.
- When the pH and temperature sensor measurement changes, the CT value of the water also changes. To disinfect a volume of water, a minimum value of CT is needed (for a pH value and a temperature value). If the new CT value is less than the required minimum CT value, the water becomes undrinkable and the WTC sends the water back to the treatment process for recycling.

B. Data Validation and Evasion

A WTC can use different comparison techniques to validate its sensor measurements. Hence, an adversary must consider to evade such data validation methods to remain stealthy. In this work, we assume multiple verification mechanisms a WTC can use to identify bad/unusual sensor measurements. Here, we discuss some of the useful approaches the WTC can following to validate its measurements:

- If a channel is not active, there should not be any sensor measurement changes for that channel.
- An adversary can attack different pH sensors and temperature sensors based on his accessibility and the sensor security. When the forged sensor measurements arrive to the controller, the controller compares the sensor measurement with the other channels. Since all channels are processing

raw water from the same natural source, their water properties should be very similar. Using this assumption, sensors of other channels can be used to validate the data of the suspected sensor.
- For a raw water source, reference CT value is periodically computed by taking raw water samples and by examining the water. From regular samples and historic database, a WTC can estimate the expected CT value. During the water treatment process, this reference value can be used to validate the calculated CT value.

Data Validation Thresholds: Considering the above data validation techniques, we define the following thresholds:

- *CT Reference Threshold*: The difference between the computed CT value and the reference CT value must be within the CT reference threshold. If the difference is more than the CT reference threshold, the data validation mechanism will identify the anomaly and notify the operator. The reference CT value is predetermined based on periodic raw water sample testing.
- *pH Difference Threshold*: For the same raw water source, pH sensor measurement value should be very similar. Based on this assumption, in this work, we consider the pH difference threshold. For a active channel, if a pH measurement is taken, the difference between the pH value and the pH value of other sources should not be more than the pH difference threshold. If the difference is more than the threshold, the data validation mechanism will identify the anomaly and notify the authority concerned.

IV. Modeling of Impact-Aware Threat Analysis

In this section, we present the formal model corresponding to WTC^2, the threat analysis framework. The formalization uses a list of parameters, denoting various WTC features, raw water properties, and attack attributes.

A. Preliminaries

We formalize the impact-based threat analysis as a constraint satisfaction problem (CSP). A CSP is generally defined using three components, i.e., $\mathbb{X}$, $\mathbb{D}$, and $\mathbb{C}$ [22], where:

- $\mathbb{X}$ is a set of variables $\{X_1, X_2, \dots X_n\}$.
- $\mathbb{D}$ is a set of domains $\{\mathbb{D}_1, \mathbb{D}_2, \dots \mathbb{D}_n\}$, corresponding to the variables ($\mathbb{X}$).
- $\mathbb{C}$ is a set of constraints $\{\mathcal{C}_1, \mathcal{C}_2, \dots \mathcal{C}_m\}$.

A domain $\mathbb{D}_i$ represents a set of values $\{x_{i,1}, x_{i,2}, \dots, x_{i,k}\}$ that are allowed for X_i. Hence, $X_i \in \mathbb{D}_i$. A constraint $\mathcal{C}_j$ is a pair $\langle \mathbb{X}_j, \mathcal{R}_j \rangle$, where $\mathbb{X}_j$ is the set of variables ($\mathbb{X}_j \subseteq \mathbb{X}$) that participate in this constraint, and $\mathcal{R}_j$ is the relation among these variables.

We solve a CSP by defining a state space. A state in a CSP represents an assignment of values to some or all of the variables, $\{X_i = x_{i,i'}, X_j = x_{j,j'}, \dots\}$. An assignment is *consistent* if it satisfies all the constraints. When every variable gets an assigned value, the assignment is *complete*. A *solution* to a CSP is a consistent and complete assignment. Here, we apply SMT logic formulas to encode and solve our proposed trajectory planning CSP. SMT provides different first order logic-based background theories (e.g., arithmetic, uninterpreted functions, bit-vectors, etc.) to solve decision problems efficiently [4], [23]. Indeed, we use Z3, an efficient SMT solver, to solve our CSP [24], [25].

TABLE I
Modeling Parameters

Notation	Definition	Type
n	Number of water channels in the treatment center	Integer
c_i	Channel i of the water treatment center	Integer
f_i	Filtration facility of the channel c_i	Integer
p_i	pH adjustment facility of channel c_i	Integer
ch_i	Chlorination facility of the channel c_i	Integer
cl_i	Original concentration of chlorine in the water at c_i	Real
ph_i	The original pH value of the water at c_i	Real
CT_i	Real-time CT value calculated from the sensors for the water at c_i	Real
$CT_{ref,t,ph,cl}$	Reference CT value for the water source for temperature t, pH ph, and chlorine cl	Real
r_{CT}	CT Reference threshold	Real
e_{CT}	CT Disinfection threshold	Real

In the following, we define different parameters and present the constraints associated with the threat analysis. The domain of a parameter is often specified by its data types (e.g., integer, real, etc.). Some parameters are constant (given as inputs).

B. Parameters

We present some of the important parameters in Table I. It is worth mentioning that no multiplication of two parameters is performed in this modeling without the multiplication sign.

In a WTC, there are n channels. Parameter c_i denotes the ith channel. Each channel is independent and contains its own water treatment processes. In each channel, there is a raw water filtration facility (f_i), a pH adjustment facility (p_i), and a chlorination facility (ch_i). Let the initial value of chlorine, temperature, and pH of water is cl_i, t_i, and ph_i, respectively. Parameter $CT_{ref,t,ph,cl}$ represents the reference CT value for the water source for temperature t, ph value, and chlorine concentration cl. Parameter L represents the required disinfection Log level. The value of L is generally fixed and predetermined based on the condition of the raw water source. In this model, no multiplication of two parameters is performed without the multiplication sign.

C. Attack Constraints

The WTC is treating raw water to make it drinkable. Let parameter V denote the total volume of raw water. Each channel has its own raw water treatment capability. Parameter v_i denote the volume of water channel c_i that can be processed on a treatment slot. If channel is active, at each treatment slot, channel c_i will pump v_i amount of water from the main raw water source. We can represent this as follows:

$$ch_{i,active} \rightarrow \hat{V} = (\tilde{V} - v_i) \tag{2}$$

Here, parameters $\tilde{V}$ and $\hat{V}$ represents the original volume of the raw water tank and the new volume of the raw water tank.

Let parameter $ch_{i,active}$ represents whether the channel is active or not. If the channel is not active (closed due to maintenance), an adversary cannot launch an attack on that channel's sensors. An adversary can attack a channel if the channel is active and attacking the channel does not violate the "minimum wait time" constraint. In this work, we denote the constraint minimum wait time using a threshold value $e_{i,w}$. If a channel is already attacked, an adversary must wait at least $e_{i,w}$ slots before it can attack the channel again. If an adversary keeps on attacking the same channel again and again without maintaining the $e_{i,w}$ requirement, the attack will become visible, and the attack will no longer remain stealthy. Let parameter a_{c_i} represent whether the channel c_i is ever attacked by the adversary and parameter $a_{i,attack}$ represent whether an adversary has the ability at attack channel c_i. We can formalize the channel attack condition as follows:

$$a_{i,attack} \rightarrow ch_{i,active} \wedge (\neg a_{c_i} \vee e_{i,w} = 0) \tag{3}$$

D. Cyberattack on the pH Sensors

During the water treatment process, the pH value of the raw water is adjusted due to two reasons. At first, the acidic or alkaline property of the raw water can be damaging for the pipes. Apart from this, the high pH value of the water makes the disinfection process less effective. Before the chlorination process, the pH value of the water is adjusted. After the pH adjustment, the new pH value of water in channel c_i is $ph_{i,new}$. We can formalize the value of $ph_{i,new}$ as follows:

$$ph_{i,new} = (ph_i + \Delta ph_i) \tag{4}$$

Here, the parameter Δph_i represents the pH adjustment of the treatment process of channel c_i.

A pH sensor measures the initial pH value, adds an adjustment and creates the new pH value $ph_{i,new}$. However, due to cyberattack, the pH measurement of the sensor changes. If parameter $\bar{ph}_{i,new}$ represents the new computed pH value of the water after the UFDI attack, we have the following constraints:

$$\bar{ph}_{i,new} = (ph_i + \Delta ph_{i,attack}) + \Delta ph_i \tag{5}$$

$$(\Delta ph_{i,attack} \neq 0) \rightarrow a_{i,attack} \wedge \neg s_{ph,i} \tag{6}$$

Here, parameter $s_{ph,i}$ represents a Boolean variable that denotes whether the pH sensor is secured. $\Delta ph_{i,attack}$ is the modification of the pH value for channel c_i.

E. Cyberattack on the Temperature Sensors

For temperature sensor, we can formalize attack equations similarly to the pH sensors. If t_i is the initial value of the temperature sensor for water treatment slot in channel c_i then the new temperature measurement of the sensor due to the UFDI attack can be represented using the following equation:

$$\bar{t}_{i,new} = (t_i + \Delta t_{i,attack}) \tag{7}$$

$$(\Delta t_{i,attack} \neq 0) \rightarrow a_{i,attack} \wedge \neg s_{t,i} \tag{8}$$

In the above equations, parameter $\Delta t_{i,attack}$ is the modification of the temperature sensor, parameter $\bar{t}_{i,new}$ is the new temperature sensor value, and parameter $s_{t,i}$ represents whether the temperature sensor of channel c_i is secured. An adversary can only alter a temperature sensor measurement, if the channel is active and attacking the sensor satisfies Equation 3.

F. Attack on the CT Value

If the chlorine residual value is cl_i, the original CT value of the water can be calculated using the following equation:

$$CT_i = .2828 \times ({ph_{i,new}}^{2.69}) \times ({cl_i}^{.15}) \times (.933^{(t_i - 5)}) \times L \tag{9}$$

Here CT_i is the original product of free chlorine residual and the contact-time.

However, due to cyberattack the pH level and the temperature of the water have changed. We can compute the new CT value using the equation below:

$$CT_{i,new} = .2828 \times \bar{ph}_{i,new}^{2.69} \times {cl_i}^{.15} \times .933^{(\bar{t}_{i,new} - 5)} \times L \tag{10}$$

Here parameter $CT_{i,new}$ is the new CT value after the UFDI attack. Let, parameter r_{CT} represent the CT reference threshold. If parameter $CT_{ref,t,ph,cl}$ is the reference CT value for the raw water source then we can formalize the CT reference threshold constraint using the following equation:

$$CT_{ref,t,ph,cl} - CT_{i,new} \leq r_{CT} \tag{11}$$

G. Clean Water Production and Recycling Constraints

If the new CT value is less than the required value, the water will not be disinfected properly. Let parameter e_{CT} represents the CT disinfection threshold. If the new CT value is smaller than the required CT value and their difference is more than e_{CT}, the water remains undrinkable and is discarded for recycling. We can formalize this as follows:

$$D_{CT} \rightarrow (CT_{i,new} < CT_i) \wedge ((CT_i - CT_{i,new}) > e_{CT}) \tag{12}$$

Here, D_{CT} is the Boolean variable that represents whether the water is discarded or accepted. If the water is discarded, it is recycled and added back to the raw water source V. If the water is disinfected, it is added to the clean water supply. If parameter $\bar{V}$ represents the new volume of raw untreated water, we can formalize this condition as follows:

$$D_{CT} \rightarrow (\hat{V} = \tilde{V} + v_i) \tag{13}$$

V. Illustrative Examples

We provide illustrative examples to discuss the outcomes of our proposed WTC^2.

A. Implementation

We implement the formal model by encoding the proposed formal model into SMT formulas [23]. In this encoding purpose, we use Z3, an efficient SMT solver [25]. The solver checks the verification constraints and provides a satisfiable (SAT) result if all the constraints are satisfied. The SAT result provides an instance that represents the value assignments to the (unspecified) parameters of the model. The attack vectors

TABLE II
EXAMPLE INPUT

```
# Volume of raw water
1000
# Number of Channels
3
# Channel number, raw water capacity, is channel operational (1-true, 0- false)?, is pH sensor secured?, is temp. sensor secured?
1 30 1 0 0
2 40 1 0 0
3 30 1 0 0
# Adversary's Resource(%)
5
# Minimum wait time
3
# CT disinfection threshold(%) and CT reference threshold(%)
20 20
```

are found from the assignments to the following variables: (i) the decision variable referring to whether channel C_i is attacked, $a_{i,attack}$ and (ii) the variables denoting the false data injections to sensors, $\Delta ph_{i,attack}$ and $\Delta t_{i,attack}$.

B. Preliminaries on the Example Scenarios

The WTC model we have used has three independent channels. We provide two examples. Our first example discusses the execution of the water treatment model without any cyberattack. In the second example, we provide a water treatment model where an adversary is launching a UFDI attack. For both of the examples, we use synthetic model data. Partial input of the example is shown in Table II.

An adversary's objective is to launch UFDI attacks on the pH and temperature sensors and remain undetected from the data validation process. Each channel is treating its share of raw water at its slot. In this water treatment facility, the raw water source contains 1000 liters of untreated water. There are three channels, each equipped with its own set of water filtration, disinfection facilities. Channel 1, 2, and 3 have 30, 40, and 30 units of water treatment capacity. Channels 1, 2, and 3 require 30 min, 40 min, and 30 min time to process 30 units, 40 units, and 30 units of water, respectively. Here, we use a random pH value and temperature of the raw water. An adversary can attack maximum five executions slots during his entire attack. In this example, we have considered the following control specifications and attack limitations:

- If the pH value of the treated water exceeds the drinking water pH range (6 - 8.5), the water is sent for recycling. Similarly, if the CT value of the final treated water is 20% less than the recommended CT value, the water is discarded and sent for recycling.
- An adversary cannot randomly change the pH measurements. Change in pH sensor measurement cannot be too different than the pH sensors of other channels. This requirement is important to ensure that the data validation process does not detect any unusual pH value in the water treatment process. In this work, we acknowledge this restriction by including the limitation that the initial pH value of the water of a channel cannot be more than 1 pH different than the average pH value of all channels.
- After attacking a channel, an adversary must wait minimum three timeslots before it can attack the same channel again.

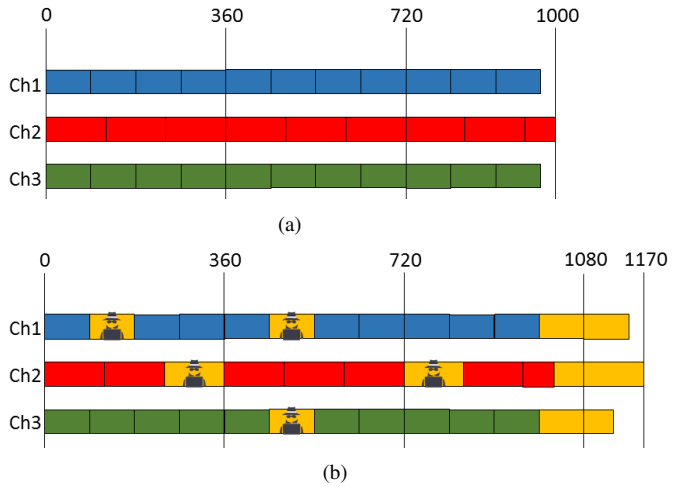


Fig. 3. (a) The execution of the framework without any cyberattack and (b) the execution of the framework with cyberattacks.

C. Execution Scenario - No Attack

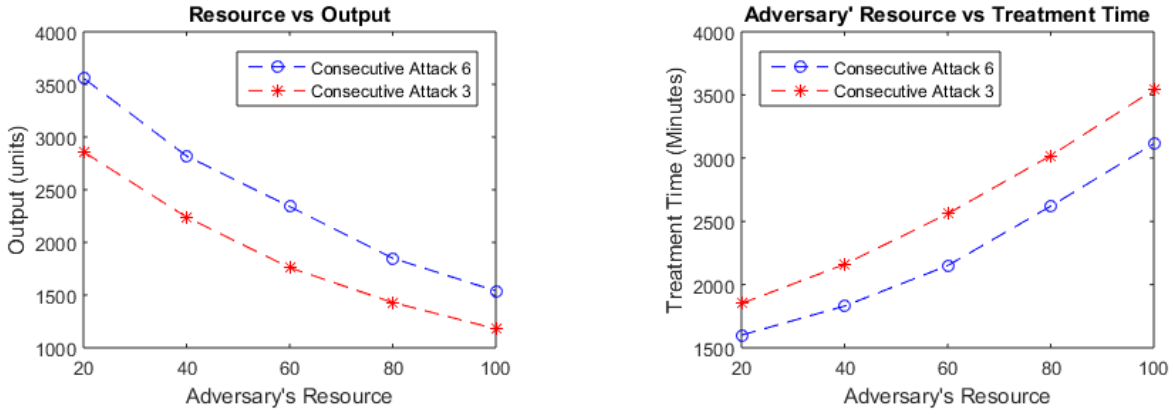
We can observe the execution of the framework in a no attack scenario in Fig 3(a). When the treatment starts, each channel pumps raw water to its own raw water tank and start the treatment process. Channel 1, 2, and 3 pump 30 units, 40 units, and 30 units of raw water from the main water tank. Once the treatment is finished, each channel pumps raw water again from the main water tank and resume the treatment process. At 30 min, channel 1 and 3 are done processing their first 30 units of raw water. Channel 2 is still processing its water. At this moment, channel 1 and 3 pump new water from the main water tank and resume the treatment process.

At 320 min, both channels 1 and 3 are processing their share of raw water. Channel 2 is available and tries to pump more water from the main water source. However, the main water source has only 20 units of water left. As a result, channel 2 pumps only 20 units of raw water and at 340 min, the execution of the framework finishes.

D. Execution Scenario - Cyberattack

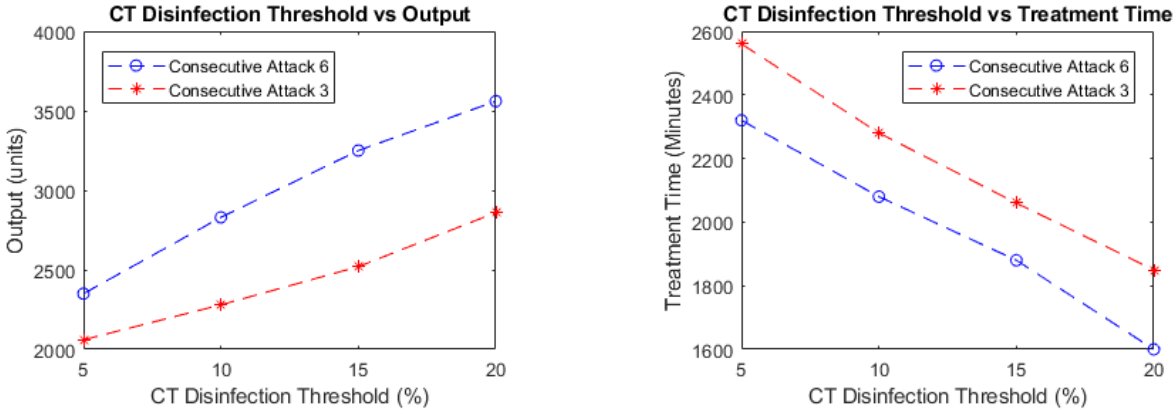
In Fig 3(b), we can observe the execution of the framework in an attack scenario. When attacks are allowed, the execution of the model returns a SAT (Satisfiable) result, along with the following variable assignments:

- An adversary attacks slot numbers 2 and 6 of channel 1, slots 3 and 7 of channel 2, and slot 6 of channel 3. The new treatment time of the facility is 400 min, which is 60 min slower than the no-attack scenario.
- In slot 2 and 6 of channel 1, both the pH and the temperature sensor is compromised. In slot 3 of channel 2, only the pH sensor is compromised. In slot 7 of channel 2, both the pH and the temperature sensor are compromised. Finally, in



(a) (b)

Fig. 4. (a) The adversary's resource with respect to (a) the water treatment center's output and (b) the total treatment time.



(a) (b)

Fig. 5. The CT disinfection threshold value with respect to (a) the treatment center's output and (b) the total treatment time.

slot 6 of channel 3, both the pH and temperature sensor is compromised by the adversary.

- Due to the attack, the WTC now need 60 minutes more to treat the same amount of raw water. Also, the attack impacted the clean water availability. At time 120 min, the water treatment facility could produce 720 units of clean water (when there is no attack). Because of the attack, the system now can produce only 590 units of water at time 120 min (18% less than the no-attack scenario).

VI. Evaluation

In this section, we present the results of the experiments that we perform to evaluate WTC^2 with respect to different attack attributes and problem sizes.

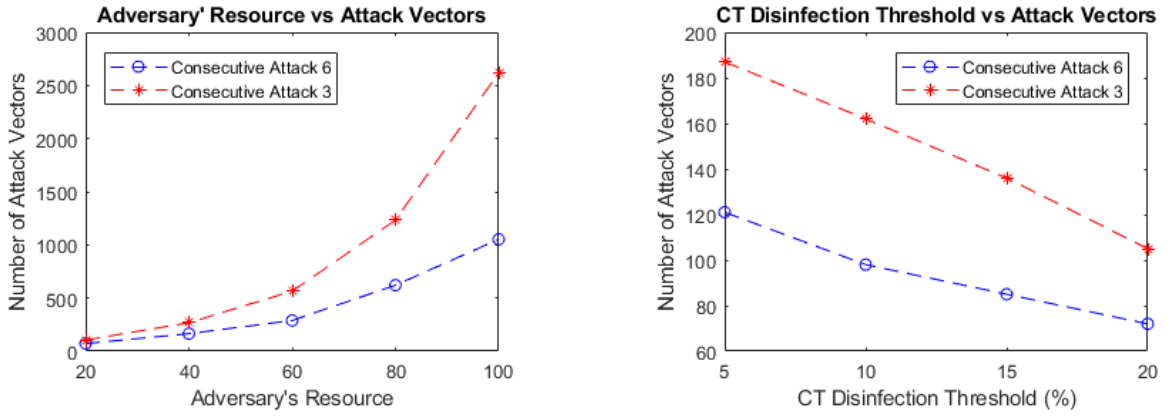
A. Methodology

We analyze the impact of UFDI attacks on the WTC, with respect to attack resource, the number of channels, and CT change threshold. We performed this analysis over two different raw water source sizes. In the scalability analysis, we analyze the execution time of our framework with respect to different number of channels, raw water source sizes, and the adversary's resource. We run our experiments on an Intel Core i7 Processor PC with 16 GB memory. We run a specific experiment several times and take the average of the results.

B. Threat Analysis Results

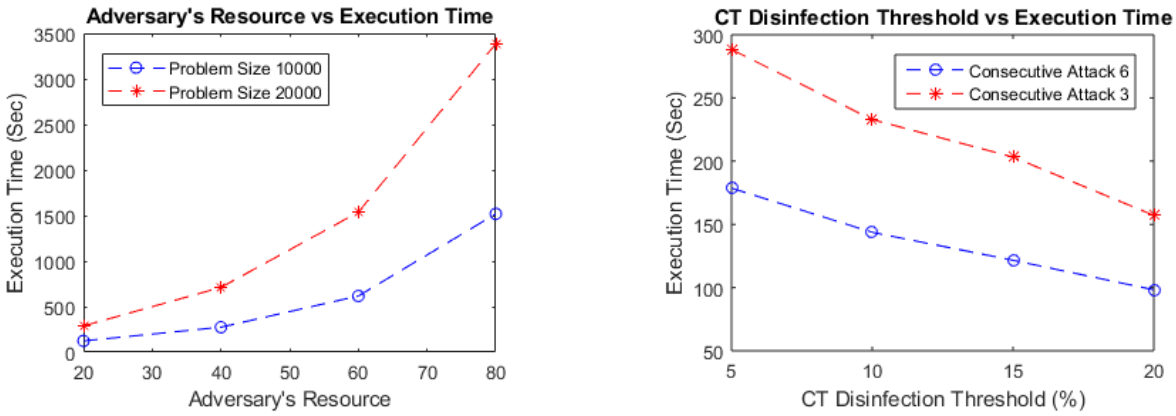
1) Impact of Adversary's Resource on Threat Analysis: From Fig 4(a), we can observe that as the adversary's resource increases, the drinkable water production capability of the WTC decreases. In this experiment, the total raw water source size is 20,000 units and there are 10 channels. In this experiment, we also compare results of two different consecutive attack thresholds. For a lower consecutive attack threshold, an adversary can attack the same channels and sensors more frequently, hence can further reducing the water production. Similarly, we can observe the relationship between the adversary's resource and the required treatment time of the WTC in Fig 4(b). From this experiment, we can observe that as the adversary's resource increases and the treatment time of the facility also increases.

2) Impact of CT Threshold Value on Threat Analysis: When the CT disinfection threshold value is small, an adversary can make small changes in the pH/temperature sensors and attack the treatment slot. As shown in Fig 5(a), as the CT disinfection threshold (in percentage) increases, the clean water output of



(a) (b)

Fig. 6. (a) the number of attack vectors w.r.t. adversary's resource and (b) the number of attack vectors w.r.t. the disinfection threshold.



(a) (b)

Fig. 7. The execution time of the framework with respect to (a) the adversary's resource and (b) the CT disinfection threshold value.

the water treatment facility also increases. For this experiment, we ran the treatment simulation for 600 min time. Similarly, in Fig 5(b), we can observe the relationship between the CT disinfection threshold percentage and the total treatment time of the facility. In this work, we assume that the raw water source contains 20,000 units of water and there are 10 channels. From this figure, we can observe that as the CT disinfection threshold increases, the problem becomes harder, the framework identifies less possible attack scenarios that violates the CT constraint, and the WTC requires less time to treat the target raw water.

3) Impact of Adversary's Resource and CT Threshold on the Number of Attack Vectors: An attack vector is an assignment of variables that satisfy the problem constraints and inflict sufficient damage. As shown in Fig 6(a), when the number of adversary's resources increases, the number of attack vectors identified by our framework also increases. In this experiment, the total volume of raw water is 20,000 units, the WTC has 10 channels, and the CT disinfection threshold value is 20%. Similarly, in Fig 6(b), we can observe the relationship between the value of CT disinfection threshold (in percentage) and the number of attack vectors identified by our framework. From the figure, we can observe that, as the CT disinfection threshold increases, the number of attack vectors identified by our framework also decreases.

C. Scalability Analysis

1) Impact of Adversary's Resource on Execution Time: As shown in Fig 7(a), when the adversary's resource increases, the execution time of the framework also increases. In this experiment, we have compared results of two problem sizes (10,000 units and 20,000 units of raw water). In the WTC model, there were 10 channels with CT disinfection threshold values of 20%. As the adversary's resource increases, our framework can identify more attack vectors and requires more time to execute. For the same adversary's resource, when the problem size is bigger, our framework spends more time for computation, hence requiring longer execution time.

2) Impact of CT Thresholds on Execution Time: We can observe the relationship between CT threshold and the execution time of the framework in Fig 7(b). When the CT increases, the number of attack vectors identified by our framework decreases, hence the execution time of the framework also decreases. In this experiment, we have compared results of two consecutive attack thresholds (threshold value 3 and 6). The WTC model is treats 20,000 units of raw water in its 10 channels. For the same CT threshold value, when the

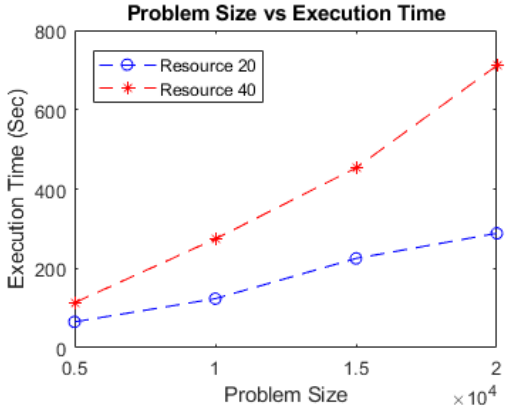


Fig. 8. The framework's execution time with respect to the problem size.

consecutive attack threshold is small, our framework takes a longer time to execute.

3) Impact of Problem Size on Execution Time: From Fig 8, we can observe the relationship between the problem size (volume of raw water) and the execution time of the framework. In this experiment, we have considered two adversary's resources (20 and 40). For both of the cases, we consider 10 channels, 20% CT disinfection threshold, and various problem sizes (5000 units, 10,000 units;, 15,000 units;, and 20,000 units; of water). From the figure, we can observe that as the problem size increases, the execution time of the framework also increases. However, even for larger problem sizes, the execution time of the framework is reasonably fast.

VII. Discussion and Conclusion

In this work, we propose a formal impact-aware framework can identify all attack vectors that satisfy the given constraints. By comparing the impact units, we can also identify the best attack vector. This framework can be very useful for off-line/on-line vulnerability analysis and can help operators to understand their system's state. Using this framework, a WTC operators can measure their resiliency against cyberattacks and can identify critical measurements. In this work, we solely focused on false data injection attacks. However, there can be other types of cyberattacks on the sensors. We can easily extend the formalization of our framework by including other attack types such as denial-of-service, jamming attack, and contention based attacks.

Accurate sensor measurements are really important for the optimal operation of the WTC. Our proposed impact-aware formal analysis framework can systematically investigate potential security threats against the WTC with respect to various system properties and attack attributes. We conduct necessary experiments to analyze the threats based on different problem sizes and constraints and to evaluate the scalability of the model. In the future, we would like to expand our work by considering other water management systems as well as experimenting on the real data.

Acknowledgement

This work is partially supported by National Science Foundation [grant number 165730/1929183].

References

[1] Charles Norman Durfor and Edith Becker. *Public water supplies of the 100 largest cities in the United States, 1962*. Number 1812. US Government Printing Office, 1964.
[2] William H Glaze. Drinking-water treatment with ozone. *Environmental science & technology*, 21(3):224–230, 1987.
[3] S. Adepu and A. Mathur. An investigation into the response of a water treatment system to cyber attacks. In *2016 IEEE 17th International Symposium on High Assurance Systems Engineering (HASE)*, pages 141–148, Jan 2016.
[4] Leonardo De Moura and Nikolaj Bjørner. Satisfiability modulo theories: An appetizer. In *Brazilian Symposium on Formal Methods*, pages 23–36. Springer, 2009.
[5] Centers for Disease Control and Prevention (CDC). Community water treatment. https://goo.gl/G8SYjN.
[6] United States Environmental Protection Agency. Chlorine - drinking water treatability database. https://goo.gl/Ftb9iG.
[7] APEC Water. Community water treatment. https://goo.gl/htSCDi.
[8] United States Environmental Protection Agency (EPA). Giardia: Drinking water fact sheet. https://goo.gl/8y2nyf.
[9] Water Research Center. C-t contact time and inactivation calculations for chlorine disinfection. https://goo.gl/2VDaSV.
[10] Yao Liu, Peng Ning, and Michael K. Reiter. False data injection attacks against state estimation in electric power grids. CCS '09. ACM, 2009.
[11] Yao Liu, Peng Ning, and Michael K. Reiter. False data injection attacks against state estimation in electric power grids. CCS '09, pages 21–32, New York, NY, USA, 2009. ACM.
[12] Gran N Ericsson. Toward a framework for managing information security for an electric power utility—cigré experiences. *IEEE transactions on power delivery*, 22(3):1461–1469, 2007.
[13] M. Amin. Security challenges for the electricity infrastructure. *Computer*, 35(4):8–10, Apr 2002.
[14] G Ericsson, O Torkilseng, G Dondossola, T Jansen, J Smith, D Holstein, A Vidrascu, and J Weiss. Security for information systems and intranets in electric power systems. *Tech. Brochure (TB)*, 317, 2007.
[15] Yacov Y Haimes, Nicholas C Matalas, James H Lambert, Bronwyn A Jackson, and James FR Fellows. Reducing vulnerability of water supply systems to attack. *Journal of Infrastructure Systems*, 4(4), 1998.
[16] Barry C Ezell, John V Farr, and Ian Wiese. Infrastructure risk analysis of municipal water distribution system. *Journal of Infrastructure Systems*, 6(3):118–122, 2000.
[17] S. Gao, L. Xie, A. Solar-Lezama, D. Serpanos, and H. Shrobe. Automated vulnerability analysis of ac state estimation under constrained false data injection in electric power systems. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 2613–2620, Dec 2015.
[18] Jill Slay and Michael Miller. Lessons learned from the maroochy water breach. *Critical infrastructure protection*, pages 73–82, 2007.
[19] Marshall Abrams and Joe Weiss. Malicious control system cyber security attack case study–maroochy water services, australia. *McLean, VA: The MITRE Corporation*, 2008.
[20] Srinivas Panguluri, William Phillips, and John Cusimano. Protecting water and wastewater infrastructure from cyber attacks. *Frontiers of Earth Science*, 5(4):406–413, 2011.
[21] E. Kang, S. Adepu, D. Jackson, and A. P. Mathur. Model-based security analysis of a water treatment system. In *2016 IEEE/ACM 2nd International Workshop on Software Engineering for Smart Cyber-Physical Systems (SEsCPS)*, pages 22–28, May 2016.
[22] Stuart J. Russell and Peter Norvig. *Artificial Intelligence*. Prentice Hall, 2010.
[23] Leonardo De Moura and Nikolaj Bjørner. Satisfiability modulo theories: introduction and applications. *Communications of the ACM*, 54(9):69–77, 2011.
[24] Leonardo De Moura and Nikolaj Bjørner. Z3: An efficient smt solver. Springer-Verlag, 2008.
[25] Z3: An efficient smt solver. In *Microsoft Research*. http://research.microsoft.com/en-us/um/redmond/projects/z3/.