A Survey of Cyber-Physical Attacks and Detection Methods in Smart Water Distribution Systems

HAJAR HAMEED ADDEEN[1], **YANG XIAO**[1], **(Fellow, IEEE), JIACHENG LI**[1], **(Student Member, IEEE), AND MOHSEN GUIZANI**[2], **(Fellow, IEEE)**

[1]Department of Computer Science, The University of Alabama, Tuscaloosa, AL 35487, USA
[2]Department of Computer Science and Engineering, Qatar University (QU), Doha, Qatar

Corresponding author: Yang Xiao (yangxiao@ieee.org)

ABSTRACT Modern technologies empower water distribution systems (WDS) for better services in the processes of water supply, storage, distribution, and recycling. They improve real-time monitoring, automating, and managing. However, the limitations of these technologies introduce cyber-physical attacks to the WDS. The main goals of cyber-physical attacks include disrupting normal operations and tampering the critical data, which have negative impacts on the WDS. Therefore, it is vital to develop and implement solutions to increase the security of the WDS by detecting and mitigating cyber-physical attacks. Since security for WDS is relatively new, there are no surveys on this topic despite its vital importance. Therefore, in this paper, we provide a comprehensive survey for the common cyber-physical attacks and common detection mechanisms for the WDS. We compare the attacks and detection methods with emphasis on ideas, methods, evaluation results, advantages, limitations, etc. We further provide a future research direction. We realize that there are still not many research attempts in this area and we hope that this work can trigger more research activities related to the WDS.

INDEX TERMS Cyber-physical attacks, artificial intelligence, water distribution systems, smart water, sensors, Internet of Things.

I. INTRODUCTION

Smart cities intend to improve the life quality of people and offer intelligent services by using smart devices. Smart cities do not have a universal definition and many scholars define them as digital or intelligent cities [1], [2]. An intelligent city collects digital data and communicates via different networks to make intelligent decisions [3]. There exist numerous components that constitute a smart or intelligent city. These components include technologies, citizens, buildings, energy systems, infrastructure, transportation systems, medical systems, government, educational systems, water systems, etc. [4], [5]. One of the most crucial components of a smart city is the water system. The sustenance of the city and the interoperability of the different components of the city is dependent on the smart water system.

A smart water system involves the use of information and technology to enhance the traditional water system. A smart water system is motivated due to the rising water scarcity and

The associate editor coordinating the review of this manuscript and approving it for publication was Yuan Tian.

the expensive structure to make water potable and available to the rising population [6]. The advantages of a smart water system over a traditional water system include accurate measurement of water consumption, quality control of water, monitoring and prevention of flooding, prevention of water wastage, etc., [7]–[11].

There are many security challenges in smart cities partially due to Internet of Things (IoT) devices to prevent water theft, water wastage, and water poisoning [14]. Also, there is a need to develop a secure water distribution system that controls the quality of water and controls the main operation water system [12]. The modern water distribution system (WDS) depends on modern technologies to operate and monitor water systems. These technologies have improved the service quality and the reliability of the WDS. However, it is susceptible to cyber-physical attacks (CPA) making them less secure. Attacks on the WDS halt or affect normal operations of these systems or even compromise critical data [13].

Cyber-physical systems and internet of things devices are considering as main components to build water distribution systems (WDS). These devices include sensors, smart meters,

programmable logic controllers (PLCs), and supervisory control and data acquisition (SCADA) systems. A sensor is responsible to collect the data of smart water, whereas smart meters measure the consumption of water. PLCs are embedded devices connected to sensors and actuators to exchange the data and control the process. While SCADA is a centralized system to manage the operations, store and analyze data [13], [15], these components increase the efficiency of WDS, but they are prone to cyber-physical attacks. An adversary can attack WDS and modify data to damage the water system or disrupt the water services [13].

Cyber-physical attacks have increased in recent years. An example attack targeted the Maroochy Water Service Sewage in Australia in 2000 [16], [17]. The adversary hacked the sewage pumping station by monitoring the radio network and sending a false message to corrupt the SCADA system and to alter the data of the sewage pumping station [16], [17]. As a result, the attack disrupted the communication for a short period between the SCADA system and the sewage pumping station and disables the operation of the sewage pump. Moreover, it prevented the alarm message to send to the SCADA system and caused the spilling out of more than 800,000 litters of raw sewage to different public areas such as parks and hotels. The government spent days cleaning up and cost around one million dollars to upgrade the security of the system [16], [17].

Another recent attack targeted an Oldsmar Water Treatment System, Florida in 2021 [18] and the hacker tried to access the computer system remotely to containment the water treatment system by changing the normal level of sodium hydroxide.

WDS can be attacked using different types of attacks such as passive attacks and/or active attacks. For passive attacks, an attacker either listens to communications among components or senses some system's states to acquire some knowledge. For example, an attacker may want to know about a given component used in the system, acquire some readings from the sensors, or listen to communications between the sensors and the actuators. Moreover, for active attacks, the attacker may make some interference to the system or components in the WDS, e.g., a) conducting Distributed Denial of Service (DDoS) attacks, b) disconnecting a component connected to the WDS, c) interfering with the communication process of the WDS, d) data breaching or e) disabling the equipment and disrupting services. Under a DDoS attack to the availability of the WDS or communications between two components in the system, the users may not access the services offered by the system [19]. Under a data breaching attack, an attacker can access data without any authority so a data breach exposes confidential and sensitive information (such as customer payment details). A data breach in the system can cause huge financial losses, reputation, and user loss of the organization [20]. Under attacks of disabling the equipment and disrupting services, an attacker can alter data pressure sensors which lead to break the pipes and loss water or change the reading data

of a water-level sensor to shut down a water pump [13]. Some active attacks are complex and difficult to detect especially when the attacker has gained many insights in the system [17].

These attacks show the security flaws inside the critical infrastructure such as water systems and there is a need to reduce the security risks or threats in the WDS. It is essential to investigate the threats and weaknesses, design detection methods, and design suitable mitigation procedures to alleviate the attacks. In recent years, there are some detection approaches to detect the attacks and mitigate the impacts of the attacks in WDS such as statistical detection methods, artificial neural network methods, machine learning algorithms, etc. [21]. Even though the area of security of WDS is relatively new and there are not many papers in this area, it is a very important area, particularly as part of a smart city. Based on our knowledge, there is no literature survey for this topic. Therefore, this motivates us to provide a comprehensive survey based on the limited and available papers. In this paper, we survey papers in detail in terms of attacks and detection methods in the WDS. Our contributions are highlighted as follows.

- To the best of our knowledge, this is the first paper to provide a comprehensive survey on security in WDS. We hope that our paper can trigger more research activities in this area.
- We survey cyber-physical attacks and their impacts on WDS. We provide a new classification of cyber-physical attacks on WDS.
- We analyze the existing detection methods to reduce the negative impacts of cyber-physical attacks on WDS.

The rest of the paper is organized as follows. Section II introduces the basic background of the WDS and components. We discuss different types of attacks for the WDS in Section III. In Section IV, we discuss existing attack detection methods for the WDS. In Section V, we present future research directions. Finally, we conclude the paper in Section VI.

II. WATER SYSTEM

A traditional water system uses fewer technologies and adopts physical devices (such as pipes, valves, and pumps) to deliver end-to-end water distribution services. Smart Water is defined as a water system enhanced with technologies such as sensing (via sensors and monitors), real-time communications (such as wireless networks, satellite communications, etc.), controls, and intelligence. A smart water distribution system is shown in Fig. 1. First, water is collected from a water source such as rivers or the sea, and then the water is transported to the water treatment plant for treating and purifying water to meet water quality standards. Furthermore, the water is stored in the water storage system and then is transported to the WDS through pipes (transmission lines), which are used to convey water from the WDS to the end-users (home, industry, etc.). Finally, after the water was used, the wastewater is transported to the water treatment
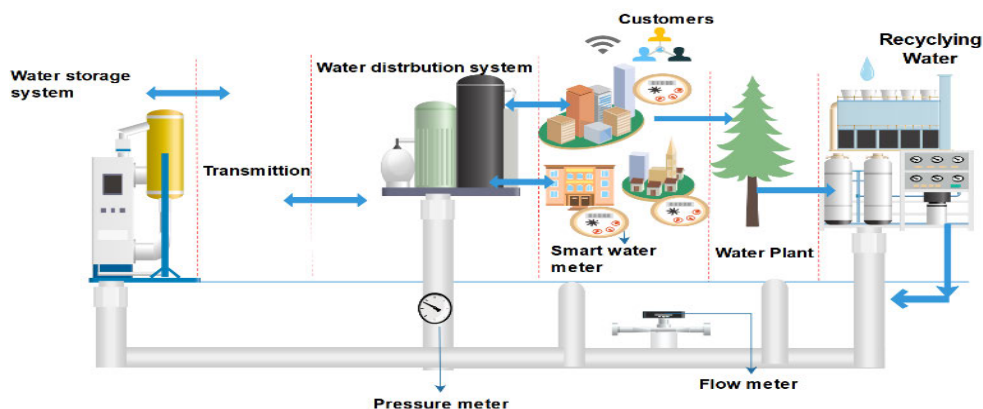
FIGURE 1. A smart water distribution system.

plant for recycling water [22], [23]. Furthermore, a smart water distribution system may include many smart devices which are physical devices with intelligent cyber features.

The major components or instruments making up a smart WDS include pipes, water tanks, smart water meters, flowmeters, smart pressure meters, energy consumption (pumping) meters, smart water treatment monitors, smart water purity sensors, physical security monitors, smart river height sensors, dam height sensors, levee movement sensors smart valves, smart pumps, smart contaminant sensors, smart flood sensors, etc. We explain some of them in detail as follows [7].

- **Smart water meters**: They are electrical instruments to measure the water consumption periodically. The water utility management collects the users' water consumption data from the smart water meters over communication networks. Then, the collected data are used for computing the cost of water consumption and managing the bills.
- **Pressure meters**: They compute the pressure of the water to sense leaking. If there is a pipe burst, flooding, or any form of irregularities, the system stops water transportation.
- **Flood sensors**: They sense the present and imminent security level for chronological and imminent flood disasters. They detect a pipe break and shut off the water supply to avoid losing water.
- **Smart valves**: They regulate or stop water streams depending on environmental situations. They allow bi-directional flows of water, disallow pressure reduction to reach the water meter, and decrease over-supply.
- **Smart pumps**: They are electronic devices to pump water through pipes and into (out from) tanks. They allow bi-directional flows of water, are used to reduce high energy usage, and can prevent a huge waste of water.
- **Smart irrigation controllers**: They contain thermostats for sprinkler systems to automatically irrigate water base on schedules or certain conditions. They can prevent loss

of water and energy that may occur due to the abuse of water usage.
- **Smart contaminant sensors**: They compute biochemical status to ensure the water quality such as temperature, turbidity, oxidation-reduction potential, pH, and conductivity. They can prevent pipe deterioration, water aging, and contaminant intrusion.
- **Pipe**: Pipes are used to transferring water.
- **Tanks**: Tanks are used to store water.

Security for the WDS becomes more important. All the above smart devices can potentially be attacked along the systems used to connect them. To ensure WDS security, it is important to design better methods to detect attacks and mitigate security risks.

III. CLASSIFICATION OF CYBER-PHYSICAL ATTACKS ON WATER DISTRIBUTION SYSTEMS

Water is one of the highest critical resources to survive on this life. Security is the crucial subject matter to the success of WDS. The modern WDS depends on modern technology such as IoT devices to perform the functions and manage the water system. Modern technology enhances the water service and increases the quality of WDS [24]. However, the limitations of IoT resources reduce the security of WDS and attract attackers to disrupt normal operations and tamper with the critical data of the water system. For instance, an adversary can attack physical components of the water supply to shut down the power of the water pump which leads to disrupt the service. Also, the attacker can alter the chemical data of treatment the water to make people sick or die [25].

Cyber-physical attacks can be classified into many types based on attacks on physical devices and attacks on the communication/networking/system components of WDS. In this paper, we classify attacks into the following types:

- **Sensor attacks**: This kind of attack is related to attacking sensors such as altering a sensor reading for the critical data.

TABLE 1. Cyber physical attacks on WDS.

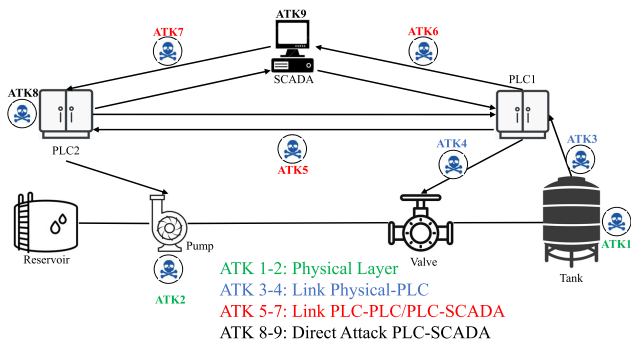| #Ref	Attack on Sensors	Attack on Actuators	Attack on SCADA	Attack on PLCs	Attack on communication links between sensors and PLCs	Attack on communication links between actuators and PLCs	Attack on communication links among multiple PLCs	Attack on communication links between PLCs and SCADA
[13]	✓	✓	✓	✓	✓	✓	✓	✓
[26]	✓							
[27]	✓							
[28]		✓						
[29]	✓							
[30]	✓	✓					✓	✓
[34]	✓	✓		✓	✓		✓	✓
[35]	✓	✓						
[38]	✓	✓						

FIGURE 2. Types of attacks in the attack model [13].

- **Actuator attacks**: this kind of attack is related to attacking actuators who make wrong actions such as making a pump more or less water to disrupt services.
- **Control system attacks**: This kind of attack is related to the control system, particularly the SCADA system which may damage whole the system.
- **PLC arracks**: this kind of attack is related to attacking PLCs, such as misleading system operations.
- **Communication/Networking attacks**: this kind of attack is related to attacking communication and networking systems so that the system can be disrupted or damaged. Examples of communications include communications among sensors, actuators, PLCs, SCADA, and other hardware and software components.

We provide a comparison among papers that introduced multiple cyber-physical attacks on WDS as shown in Table 1.

Cyber-physical attacks can largely impact the WDS. However, there is a lack of existing analytical and computational tools that characterize the responses of WDS to different kinds of cyber-physical attacks and show the risks of assaults. The authors in [13] aim to build attack models that characterize different types of cyber-physical attacks to identify the components of the cyber-physical system of WDS that respond to the attacks as shown in Fig. 2 and summarized in Table 2. Moreover, the authors in [13] build the EpanetCPA toolbox enables the researchers to design

the attacks and customize the attributes for every type of attack. Also, it can simulate the attack results of responding to WDS. The authors in [13] study the characteristics of attacks depending on duration time, e.g., the starting and ending time for a specific attack, the number of attacks that occurred, and the locations of the components that will be attacked. They perform attacks in different scenarios as follows:

- In the first scenario, the attack aims to control pump PU1 and pump PU2 at the same time and forces both pumps to run 10 hours without any demand from Tank T1, shown in Fig. 3, known as the C-town network. As result, the tank overflows 35 hours because the attack disallows PU2 to stop when the level of water of T1 is higher than thresholds (4.5 m).
- In the second scenario, the attack targets to alter the water level reading that was collected by PLC3 in T4. Then, it sends the altered information of PLC3 that controls PU6 and PU7 to give a wrong decision to deactivate pumps 6 and 7. Therefore, the water level in T4 is decreased by 1.11 hours.
- In the third scenario, the attack alters the data on the T1 water level sent by PLC2 to PLC1 leading to the overflow in T1.
- In the fourth scenario, the attack aims to hack the communication link between PLC2 and SCADA system, and then to alter the data reading of T1 water level. As result, the SCADA system stores incorrect information.
- In the fifth scenario, the attack targets the communication link that connects SCADA and PLC5 and modifies the thresholds that activate/deactivate PU11. As result, PU11 is activated for 50 hours which increases the water level of T7 and makes a wrong decision for PLC9 to deactivate PU10.
- In the last scenario, 100 random attacks are used to attack PLC3, T1, T2, and T3.

As a result, the C-town water network shown in Fig. 3, is largely impacted during the periods of attacks [13].

Furthermore, the authors in [27] propose Man-in-the-Middle attacks to attack a water treatment testbed, called

FIGURE 3. C-town water distribution network [30].

SWaT, which contains physical devices such as sensors, actuators, and PLCs. When the sensors collect data and send the data over a Fieldbus communication network to a PLC, an attacker can spoof packets through the Fieldbus communication protocol and inject a false date to modify the sensor reading [27]. As a result, the attack successfully accesses the physical devices and controls the whole SWaT testbed [27].

The authors in [26] present attacks to a water distribution testbed (WADI) which uses to purify the water. WADI is prone to attacks because its components such as sensors, PLCs, and Remote Terminal Units (RTUs) are subject to attacks. WADI also is connected with SWaT to supply filtered water. They design two types of attacks and show that attacks could cascade to the SWaT. The first attack alters the sensor reading to reduce the level of water by changing the percentage of the level water from 75% to 10% to turn on a pump and this makes it overflow. The second attack alters the sensor reading of a tank to shut off the tank valve. The results show that the attacks successfully impact the WADI, cascade the failure to SWaT, and impact the operations. However, the authors in [26] only list two types of attacks and do not provide a detection mechanism to detect these attacks.

An American water utility company was attacked by hacking the passwords of the routers to access and monitor the pumping stations in 2016 [29]. As a result, the monthly bill price was increased from $300 to $15,000 and later the utility company upgraded the security to mitigate attacks [29].

The authors in [1] launched an availability attack and an integrity attack on a water system testbed developed by an EU project called ''FACIES'' which has a fault diagnosis module to detect anomalies. The results show that the security of the system is poor and unable to detect the launched attacks. Therefore, the fault diagnosis module needs to be enhanced with a robust detection method to increase the security of the water system testbed.

IV. DETECTION METHOD CLASSIFICATION OF WATER DISTRIBUTION SYSTEMS

The purpose of a WDS is to deliver better water service to consumers. However, WDS is prone to cyber-physical attacks which can disrupt normal operations and tamper with the critical data of the water system. The attacks expose the main components of the water system. For instance, the advertisers can modify the values of water levels in each tank or the suction pressure in each pump station to

TABLE 2. Cyber-physical attacks in water distribution system [13].

Attack#	Objective	Function	Consequences of the attack
1	Directly access to the sensors such as water level sensors in tank water.	Change, manipulate, or replace data for damage, e.g., altering data of water level sensors to manipulate control operations of turning on/off the water pump.	Disrupt service.
2	Directly access to the actuators such as pumps or valves.	Alter data, e.g., changing the data related to the pumping speed or implementing DOS attacks.	Disrupt availability of the service and compromising of information integrity, e.g., changing the data of pumping speed can affect on pushing and delivering water.
3	Attack the communication links between two components which are a sensor and a PLC.	Alter data, e.g., changing collected data package in PLC, discovering the information of the statue water system, or launching DOS attacks.	Disclose the sensitive information of the system and disrupt the service.
4	Attack the communication links between two components which are an actuator and a PLC.	Eavesdrop the control signals between an actuator and a PLC or alter the signals to give incorrect information.	Damage a water system.
5	Attack on the communication link that connects between multiple PLCs, e.g., PLC1 gathers data from sensor water level in tank water and transmits to another PLC2 which controls the pump operations in another tank water.	Eavesdrop or alter data of PLC and gain access the SCADA system.	Control pumping operations and interrupt the service.
6	Attack on communication link between two components which are PLCs and the SCADA system.	Eavesdrop data of PLC, interrupt the service by flooding traffic to the communication channel and send wrong information.	Gain access the SCADA system and produce wrong decisions to control operation.
7	Attack on the communication link between two components which are the SCADA system and PLCs.	Prevent or hinder the SCADA system to send signal control data to PLCs or prevent to activate/deactivate pumping of water.	PLC cannot receive signal control data or receive wrong information which can affect on scheduling pumping operations.
8	Attack on PLC.	Fully control on PLC and implement DOS attacks.	Make PLC completely out of work and reach high level of controlling.
9	Attack on SCADA.	Chang the configuration data of water system or alter collected data of sensors.	Destroy a water system completely.

turn ON/Off the pump water. Also, the adversary can alter flow sensors values in the valves or pump stations to causes physical damages. It is important to develop an anomaly detection algorithm to identify the anomalies before they harm the system. The development algorithm also should identify the global anomalies from multiple sensors in a multi-dimensional space. Note that in contrast to a global anomaly, a local anomaly indicates a case that from a single sensor. A detection algorithm that only recognizes local anomalies from each sensor separately might miss the attacks due to the potential high dimensionality of the sensor's data. Also, there is a probability occur a single attack that can affect multiple sensors at the same time and mislead the operations of the system [30]. We will survey many detection methods first and then compare them in the following subsections.

A. DETECTION METHODS
In this subsection, we survey most of the detection methods in the literature in detail as follows.

The major contribution of the authors in [13] is that they propose nine types of attacks to the components of the

water distribution system such as sensors, PLC units, and SCADA system as shown in Fig. 2. The detection method is simple and trivial. For the limitations, the attacks model is applied on a simple WDS that has one pump, one tank, one valve, and a few actuators; experiments on a complex water distribution system network are needed as well as a well-designed detection method.

The authors in [30] develop a detection algorithm to identify local anomalies that affect each sensor individually and also identify the global anomalies that affect multiple sensors at the same time. The proposed algorithm includes three layers as shown in Fig. 4: 1) a simple statistical detection layer to determine outliers; 2) an Artificial Neural Network Model (ANN) layer to detect contextual anomalies based on data from one sensor, and 3) a Principle Component Analysis (PCA) layer to detect anomaly behaviors among multiple sensors. We explain the detailed methods, attacks, results, and limitations as follows.

- *Layer-1*: For the first layer, a simple statistical detection method to determine outliers by calculating the upper and lower boundaries. The upper boundary is equal
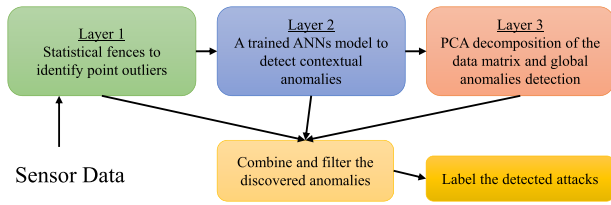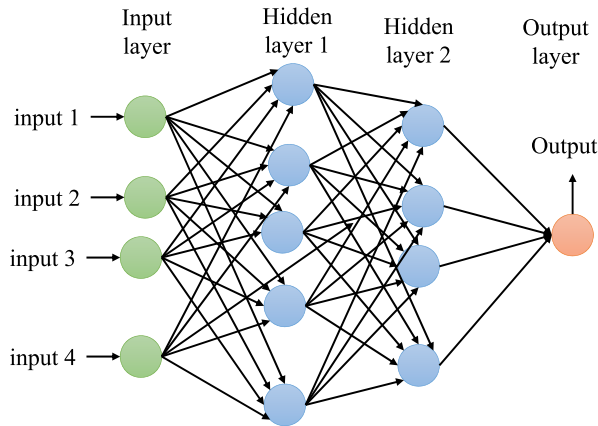
FIGURE 4. An anomaly detecting algorithm [30].

FIGURE 5. An artificial neural network model [31].

FIGURE 6. The mathematical of artificial neural network model [32].

PC will classify the dataset as normal for the dataset whose maximum variance is large; otherwise, the PC will classify it as an anomaly for the lower variance dataset.

- *Attacks*: The authors also adopt the C-Town Distribution System as shown in Fig. 3. The authors adopt two datasets: 1) the sensor data without attacks over one year and 2) the dataset with labeled five kinds of attacks over 6 months. As shown in Fig. 3, the first attack targets to attack PLC5 that connects between PU10 and PU11 and the connection between PLC9 and PLC5. The second attack targets to attack water level data of tank1 which affects on control operation of PU2. The third attack targets to change the pressure value of the valve and pumps 4 and 5. The fourth attack intends to attack PU7 by attacking the water level data of Tank4. The fifth attack targets to attack PU6 and PU7.

- *Results*: The algorithm is tested on C-Town Distribution System as shown in Fig. 3. The proposed algorithm can identify all the labeled attacks with no delay. Also, it identifies three new attacks that have duration attacks less than the original minimum expected attacks duration in the given compromised set.

- *Limitations*: The limitations are summarized as follows. First, the third layer in the algorithm that adopts PCA could not recognize the exact component that was compromised during the detected attack 3 because the discovered anomalies are obtained from multiple components at the same time. Also, the authors do not provide a solution for this so that there is a probability to lose a huge amount of water in case that the physical attacks damage assets of the water system without identifying those compromised components. Second, the algorithm issues false alarms which put the system under attack status for a few hours even after there is no attack existed. Third, training the ANN with increasing the number of hidden layers is time-consuming.

to the sum of the mean value and the product of the standard deviation and an upper multiplier in the data set, while the lower boundary is equal to the subtraction of the mean value from the product of the standard deviation and a lower multiplier in data set. Therefore, any value that is greater or less than the upper boundary or lower boundary, respectively, will be classified as outliers.

- *Layer-2*: For the second layer, a trained ANN model includes multiple layers of interconnected artificial neurons that perform nonlinear computations as shown in Fig. 5. It has an input layer, multiple hidden layers, and an output layer. The input layer consists of independent neurons and each neuron is multiplied by a connection weight to the hidden layers. Then, the combination of bias and all weighted inputs is fed to an activation function as shown in Fig. 6. If the result of the summation process is greater than a threshold of the activation function, the neuron output will be able to be fed to the next input layer. To adjust the relationship between the input layer and the output layer, every neuron has a specific weight which can be adjusted by applying a Backpropagation (BP) algorithm to update the weight. The BP algorithm is used to compute a gradient and to reduce errors in the predicted output values of the training data set [33].

- *Layer-3*: For the third layer, the PCA method is used to reduce the multi-dimensional sensor data and to project it to new axes called Principle Components (PCs). Each
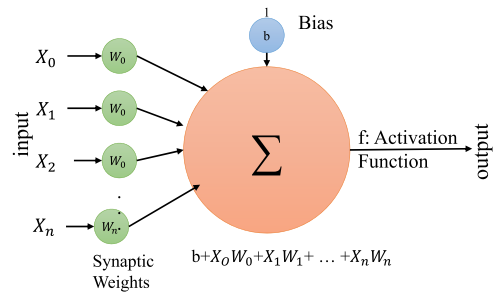
The authors in [34] propose an algorithm to identify malicious attacks by checking the integrity of sensors' data and actuators' rules and identifying the anomalies in the data. Then, they adopt an optimization approach to extract low dimensionality of sensor data and separate them from

all the SCADA data measurements. We explain the detailed methods, attacks, results, and limitations as follows.

- *Part I*: An attacker can attack the water level sensors to change the values. The first part checks the integrity of the data system and actuators' rules to protect the SCADA system from compromising by attackers. The first part includes two functions.

  - - The first function is called "Actuators Rules Verification" which is to check the integrity of the actuators' rules. For example, one of the basic rules of actuators is to activate or deactivate a pump based on the measurement of water level sensors. If the water level in a water tank is less than the minimum threshold, the pump should be activated. Otherwise, the pump should be deactivated. If this rule is not followed, there is an attack on the sensor data or actuators' rules. The algorithm also determines the time of the attack.

  - - The second function is called "Data Verification" which is to compare the data of all variables with the range of the normal operation. First, the function determines the minimum and maximum water levels values and then checks if any level reading is less or greater than the bounds of the minimum and maximum water level values. Second, the function checks the pump statue with the pump water flow value. For instance, if the pump statue is turned ON, the pump water flow value must not equal to zero; otherwise, the pump water flow must equal to zero. Third, the function also checks if the valve statue works correctly with the flow valve. Finally, the function checks the level of pressure to see if it is below or above normal threshold values of pressure.

- *Part II*: The second part detects anomalies and isolates them from the actual SCADA system data set based on an optimization algorithm. The measurement values of the SCADA system during the period can be represented as a matrix. The matrix includes the true values of SCADA measurements without interrupting by noise or attacks (L), false values of attacks (A), and noise values (N). The total measurements data is donated by (Y) at a specific time as Y = L+A+N. The authors assume that L is a low-rank matrix and it is nonconvex so that it is difficult to minimize it computationally. To convert the matrix to a convex optimization problem and to extract the low dimensional data from SCADA, they run an optimization framework called CVX. CVX is software that uses Matlab language to model the convex optimization problem. It requires a training data set which is Y measurements of the SCADA system. Also, the software selects tuning parameters to minimize the percentage of errors in calculating the convex problem to extract the anomalies. Moreover, the authors assume that A is a sparse matrix to determine the attacks. If A = 0,

there are no attacks on the measurements values (Y). The A matrix should contain zero values in all the rows and columns except one row that can have non-zero values. Then, they determine the threshold value of A and compare the non-zero values with the threshold. If the non-zero values are greater than the threshold, the attacks are presented.

- *Attacks*: The authors also adopt the C-Town Distribution System as shown in Fig. 3 under eight attacks. The first attack targets to attack the low-level water sensor in Tank5. The second attack targets to attack a high-level water sensor. The third attack targets to attack a water overflow sensor in Tank1. The fourth attack tends to attack PU10 and PU11. The fifth and sixth attacks target actuator rules of controlling PU2, respectively. The seventh attack targets to pressure value of PU7. The eight attack targets actuator rules of controlling PU6 and PU7.

- *Results*: The authors test the algorithm on the C-Town Distribution System as shown in Fig. 3 under eight attacks. The results show that the data verification algorithm can detect all types of attacks except the seventh and eighth types of attacks, while the actuator rule algorithm identifies all labeled attacks for compromised data set.

- *Limitations*: The limitations are summarizing as follows. First, the Data Verification algorithm could not identify all the eight labeled attacks. Second, the low dimensional data still appear on the SCADA measurements although the purpose of adopting the optimization algorithm is to extract low dimensionality sensors' data completely from SCADA measurements' data to detect attacks. In other words, the optimization algorithm is not very effective.

The authors in [35] propose an intrusion detection method to detect Denial of Service attacks (DDoS) in smart cities. DDoS attacks can disrupt the service by flooding the system with large numbers of requests and make it inaccessible to the users [37]. For instance, DDoS can affect the performance of a smart water plant by sending large numbers of requests to a data controller as shown in Fig. 7. As a result, a data controller sends harmful control signals to the water pumps causing the change of the pumping speed at the pumps and the disruption of the services. A machine learning algorithm is applied to the smart water plant dataset to detect DDoS attacks that disrupt the services by changing the pump speeds. The algorithm clusters DDoS attacks to multiple attack types, build a Deep Restricted Boltzmann Machine model (RBM) and build a Feed-Forward Neural Network model (FFNN) to detect the attacks. We explain the detailed methods, attacks, results, and limitations as follows.

- *Step-0*: The initial step is to determine each type of DDoS attack based on the dataset by applying a K-mean clustering algorithm. It defines k categories of data and identifies for each category centroid point. Then,
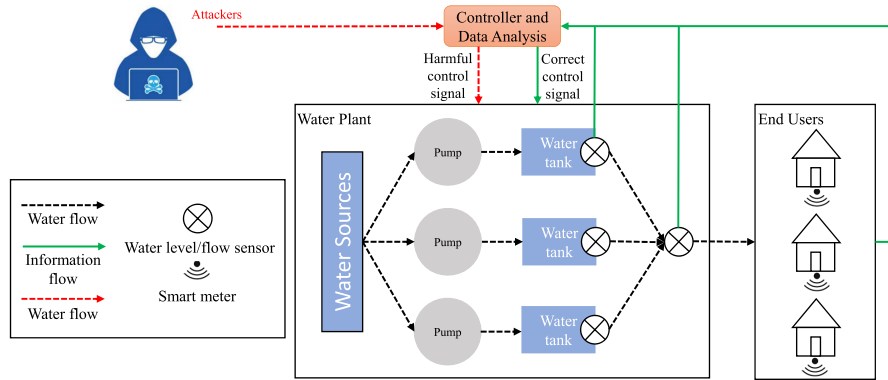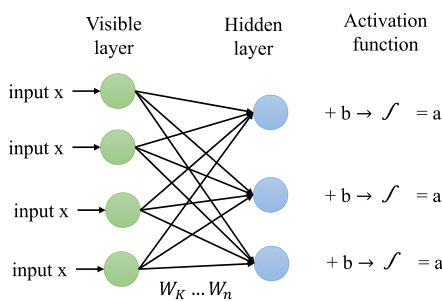
FIGURE 7. DDoS attacks on smart water plant [35].

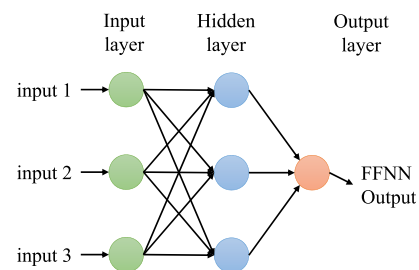**FIGURE 8.** Restricted boltzmann machine model [36].

FIGURE 9. Feed forward neural network model [37].

it calculates the distance between the elements and the centroid point and assigns the elements to their categories based on the smallest data centroid.

- *Step-1* The first step is to apply the RBM model that learns the high-level six features of smart water plant data from sensors and smart meters. These features include water level 1, water level 2, water flow 1, water flow 2, pump speed 1, and pump speed 2. The RBM model contains one input layer and multiple hidden layers and there is no output layer as shown in Fig. 8. Each input layer has many variables multiplied by a specific weight which are then connected to the hidden layers. The summation process of all products with a bias is fed to an activation function to produce the output that determines if the hidden state is activated or not. Then, the reconstruction is calculated in the same way, but in the opposite direction that starts from the hidden layers to the input layers.

- *Step-2*: The second step is to build FFNN model layers that consist of one input layer with multiple neurons, one or more hidden layers, and one output layer as shown Fig. 9. Each neuron in the input layer has a directed forward connection to the neurons in the hidden layers without any loops or cycles. The FFNN model is trained to classify smart water plant data as normal data or anomalies as DDoS attacks that change the speed of the water pump.

- *Results*: The results show that the performance of the FFNN model algorithm without applying the RBM model layers achieves an accuracy rate of 93% and the FFNN model algorithm with the RBM model including only one layer achieves a high accuracy of 97.5%. However, the accuracy rate is dropped when the number of RBM model layers increases to 97.1%. Also, the FFNN model algorithm trained with the RBM model including two layers achieves an accuracy rate of around 97%, while the FFNN model algorithm with the RBM model including three layers achieves an accuracy around 97%.

- *Limitations*: The limitations are summarizing as follows. First, the increase in the number of layers in the RBM model hurts the performance of the system and the RBM model becomes unable to detect the attacks. Second, when the RBM model converts sensor data of smart water plants to binary representation, it causes loss of the information and reduces the reliability of the performance. Furthermore, the authors only provide a detection method without any recovery solutions to recover attacks and isolate the compromised water components.

The authors in [38] propose an algorithm to identify anomalies including three modules: control rule and consistency module, pattern recognition module, hydraulic and system relationships module. We explain the detailed methods, attacks, results, and limitations as follows.

- *Control rule and consistency module*: The control rule and consistency module check the consistency of data with specific control rules given in the data set of control rules. Based on the control rules, PLCs control the water system and give control commands based on the collected data from sensors and actuators. An adversary can tamper with the control rules and also modify the obtained data to cause some inconsistency in the WDS dataset.
- *Pattern recognition module*: The pattern recognition module contains distinct patterns for hydraulic parameters. There is a pattern for every hydraulic parameter and also there is a pattern for a combination of multiple hydraulics parameters. These patterns are developed based on datasets free from cyber-attacks. Then, the authors compare these developed patterns with the current WDS dataset to identify the anomalies. If the WDS dataset does not follow the developed patterns, anomalies occur.
- *Hydraulic and system relationships module*: The hydraulic and system relationships' modules are developed based on the relationships of WDS components. For instance, the physical information of the water level in the tank and pump station flow can be calculated to derive the information of demand from another component such as mass balance. The calculated values of these components are compared with the collected data of WDS components to detect the attacks.
- *Results*: The authors test the algorithm using a collected dataset for 6 months. The algorithm shows the efficiency of the algorithm to identify the anomalies and detect the attacks. The total hours of operations are 4177 hours and the anomalies are observed in 666 hours. The algorithm detects 62 cyber-attacks and calculates the duration time to detect each attack. The longest period that the algorithm spends to detect the attack is 87 hours, while the shortest period to detect the attack is 2 hours.
- *Limitations*: The limitations are summarizing as follows. First, the algorithm detection time needs to be enhanced to minimize detection time. Second, comparing the water dataset with the given patterns of modules is not enough to detect some attacks.

The authors in [44] propose a methodology to detect cyber-physical attacks on WDS by molding the component of WDS as a logic graph, e.g., a graph contains interconnection between sensors and actuators. Then, the algorithm in [44] computes a control function for each subgraph and compares the computed value of the control function with the new value of the control function. When an attacker tries to compromise the components, the new value of the function control graph will be different from the original value of the control function and the attack occurs. The algorithm also recovers the attacks by replacing the comprised value of water components with an estimated correct value to mitigate the impact of attacks. The method was tested on a simple WDS and needs to extend in the future for testing on a complex WDS.

The authors in [39] propose an ensemble methodology to detect cyber-physical attacks by creating four modules. The first module is responsible to check if the operations of pumps and valves follow the right control rule based on the observed water levels in each tank. For example, if the water level in a tank is less than the threshold, the pump should be turned on. The second module contains a statistical method to monitor the sensors and actuators based on the calculated upper and lower boundaries. If any value is above the upper boundary or below the lower boundary, it will be considered as outliers. The third module has an Artificial Neural Network Model (ANN) model to train the dataset and predicts the anomalies. The fourth module contains Principle Component Analysis (PCA) to classify data as normal or anomalies. The results show the ability of the method to detect all the labeled attacks and it identifies the locations of the compromised components for most of the labeled attacks. However, there is a need to develop the algorithm to increase the ability to recognize the compromised components when attacks occur from multiple components at the same time.

The authors in [45] propose a model-based fault detection methodology to detect cyberattacks on WDS. First, the authors create a dataset free from attacks by adopting the EPANET simulator to simulate the WDS. They define the normal errors which are expected in the WDS due to the differences between the EPANET and the actual WDS and then create another dataset with attacks to produce errors in the SCADA readings [45]. The proposed algorithm in [45] compares the normal errors and the errors produced in the presence of attacks for SCADA readings to discover potential cyberattacks. The results show the ability of the algorithm to detect attacks from SCADA readings, but the reliability of the algorithm is decreased since the sensor's data with noise affects the accuracy of the results.

The authors in [40] adopt Long Short Term Memory Recurrent Neural Network (LSTM-RNN) method to train the SWaT testbed dataset and predict the attacks. RNN is one type of deep neural network and it feeds the output layer as the input for the next layer. A propagation algorithm is used to update the weight for the gradient value to increase the ability of the algorithm to predict results. However, RNN has a vanishing gradient problem so that the algorithm cannot update the weight because the gradient value is very small and this leads to a poor result. The LSTM algorithm is used to solve the vanishing gradient problem in RNN and increases the ability of model learning. LSTM includes a memory block that consists of an input gate, a forget gate, and an output gate to increase the ability of the model to learn data and predict results. The results in [40] show that the proposed method detects the labeled attacks with high accuracy. However, the proposed method is trained with a small sensor dataset and it should be extended to train whole the sensor dataset in the SWaT testbed.

The authors in [41] adopts Autoencoder neural networks (AE) algorithm that includes an encoder and a decoder, where the encoder encodes and processes the input data to reduce dimensionality and the decoder reconstructs data from encoded data and obtains a small reconstruction error measured from all data. If the reconstruction error is larger than the small reconstruction error, an anomaly occurs. The proposed algorithm in [41] can detect all the labeled data, but the performance is decreased when all the SCADA system is compromised.

Most of the WDS adopt a single site Event Detection System (SSEDS) to detect contamination based on sensor readings. The authors in [46] develop a Multi-Site Event Detection System (MSEDS) method to detect contamination for all the datasets of WDS and not only for the sensor dataset. The proposed method simulates the quality of water-based on six characteristics: "Chlorine, Electric Conductivity, pH, Temperature, Total Organic Carbon, and Turbidity". Then, the method calculates the errors between sensor readings and the predicted values of the simulation model. Moreover, the method classifies these errors into normal or abnormal based on thresholds that determine the six water quality characteristics. The method focuses only on water contamination and improves the detection of contamination, but it does not study other attacks and detection.

The authors in [47] propose a smart fuzzing method to identify network attacks that spoof the sensor readings and command actuators over the network. They adopt two machine learning algorithms, LSTM and Support Vector Regression (SVR), to train data set and classify datasets based on thresholds of fitness functions. If the predicted value is greater than the threshold of a fitness function, an attack occurs. The results show that the proposed method can identify 27 attacks, but the performance is slow. Moreover, the method does not detect attacks other than network attacks.

The authors in [48] propose a method to detect DDoS attacks and test on a dataset of smart water distribution plants. The method adopts the RBM model to reduce the dimensionality of data and uses four different algorithms to classify data into normal and abnormal data. First, they determine 10 types of DDoS attacks based on a k-means clustering algorithm. Then, they apply the one-layer RBM model to four algorithms which are FFNN, FNNN automated, rain forest, and support vector machine to classify data. Then, they apply the two-layer RBM model to the four algorithms and repeat the same process until the RBM model reaching to 5 layers. The results show that the FNNN automated algorithm has the best accuracy to detect attacks. Furthermore, the performance is the best when the RBM model has one layer and with the increase of the number of layers in the RBM model, the complexity is increased, causing the decrease of the performance of detecting attacks.

The authors in [42] design some jamming attacks which block the communication channels to disallow the communications between PLCs and the physical process. The goal of the jamming attacks is to control or damage a SWaT testbed.

The results in [42] show that the SWaT testbed is responded to and had negative impacts on water overflow. Furthermore, the authors in [42] provide a simple detection method to detect attacks by comparing the measurement values with their properties. However, the detection method needs to be improved as future work to enhance the security of both the physical layer and the network layer.

A competition called BATADAL was organized in a conference in California in 2017 to compete for attack detection algorithms to detect cyber-physical attacks in the WDS [21]. The authors in [21] summarize various detection methods of seven team participants at BATADAL and evaluate them based on metrics such as detection time and the ability to identify the compromised components. The first team adopts a method that extracts features based on calculated mean and covariance and then adopts a rain forest algorithm to classify data into normal and abnormal data. The second team adopts district metered areas to reduce the dimensionality of data and then adopts recurrent neural networks to classify data and to predict the attacks. The third team proposes a method that first checks the integrity of rule operations, and then classifies data based on a deep neural algorithm called convolutional variational autoencoder. The fourth team proposes a method that first checks the integrity of SCADA data and actuators' rules, and then adopts an optimization algorithm to minimize the computation time. The fifth team proposes a model that has three layers to detect attacks, including a layer using an outlier detection method, a layer using an artificial neural network algorithm to classify data, and a layer to determine the anomalies based on principle component analysis. The six-team proposes a method with three modules to check control rules, integrity of data, and the relationships among components of WDS. The seventh team proposes a model based on EPENANT to simulate a WDS and compare the data of the actual water system with the simulation model to detect the attacks. The results show that all the teams can detect cyber-physical attacks, but the seventh team won the competition by achieving the best performance overall. However, all the proposed algorithms are trained based on the medium size of a WDS and the large size of WDS should be considered in the future.

B. COMPARISONS

We compare existing detection methods as shown in Table 3. We also provide a general comparison of detection methods based on different selection criteria as shown in Table 4. We compare some of them as follows.

These papers present in-depth researches on the impacts of cyber-physical attacks on water distribution systems (WDS). The objective of these studies is to detect cyber-physical attacks by identifying anomalous behaviors in the WDS. Table 1 shows the different types of cyber-physical attacks among several papers that disrupt the services and compromise the main components of WDS. The paper [13] proposes all the cyber-physical attacks while the papers [30], [34],

TABLE 3. Comparison of detection methods.

#Ref	Method	Details of the methods	Combined method
[13]	A simple method	It is an object-oriented programming where the nine attacks are implemented as classes. For each class's attributes, the simple method determines the features of attacks, such as the action of an attack and the time duration, etc.	N/A
[30]	Statistical	It calculates the upper and lower boundaries based on the mean and the standard deviation. If any value is greater than the upper boundary or less than the lower boundary, it will be classified as outliers.	First, the statistical method identifies the outliers in each sensor individually by comparing the value of data with high and low boundaries of the normal operations. However, there is the possibility to detect the false outliers. Then, the Artificial Neural Network Model (ANN) model is trained to learn the patterns of normal operations and to predict the future anomalies in each individual sensor. However, there is an overfitting problem. Finally, the Principle Component Analysis (PCA) method reduces the high dimensional from all combined sensors' data and detect anomalies from multiple sensors on the same time.
	ANN	It has an input layer, multiple hidden layers, and an output layer of a neural network. The input layer consists of independent neurons and each neuron is multiplied by a connection weight to the hidden layers. Then, the combination of all weighted inputs is fed to an activation function. If the result of the summation process is greater than a threshold of the activation function, the neuron output will be able to be fed the next input layer.	
	PCA	It reduces the multi-dimensional sensors and projects it to new axes called Principal Components (PCs). Each PC counts maximum variance data by subtracting each value from the mean value. Then, PCs can separate in two datasets. One set has a maximum variance of all data and classifies it as normal data. The other set has the lowest variance of data and classifies it as anomalies.	
[34]	Actuators Rules Verification Algorithm	It checks the integrity of the actuators' rules. For example, one of the basic rules of actuators is to activate or deactivate a pump based on the measurement of water level sensors. If the water level in a water tank is less than the minimum threshold, the pump should be activated. Otherwise, the pump should be deactivated. If this rule is not followed, there is an attack on the sensor data or actuators' rules.	Frist, the actuators rules and data algorithms check the integrity of SCADA measurements. Then, the optimization algorithm runs to detect the complex cyber-physical attacks that cannot be detected by the previous algorithms.
	Data Verification Algorithm	It compares the data of all variables with the range of the normal operation. For instance, the function checks the pump statue with the pump water flow value. If the pump statue is turned ON, the pump water flow value must not equal to zero; otherwise, the pump water flow must equal to zero.	
	Optimization Algorithm	It is used to extract and isolate low dimensional sensor data from all the SCADA data. It runs an optimization algorithm on the SCADA measurements data everyone hour by using convex optimization software such as CVX. The CVX detects the anomalies and minimizes the computational time.	
[35]	RBM model	It contains one input layer and multiple hidden layers and there is no output layer of a neural network. Each input layer has many variables multiplied by a specific weight which is then connected to the hidden layers. The summation process of all products with a bias is fed to an activation function to produce the output that determines if the hidden state is activated or not. Then, the reconstruction is calculated in the same way, but in the opposite direction that starts from the hidden layers to the input layers.	First, the RBM model is trained to learn high features from sensors' readings. However, the RBM model must convert real sensors' data to binary data which causes losing information. Then, the FFNN model classifies the data to normal data or abnormal data.
	FFNN model	It consists of one input layer with multiple neurons, one or more hidden layers, and one output layer of a neural network. Each neuron in the input layer has directed forward the connection to the neurons in the hidden layers without any loops or cycles.	
[38]	Control rule and consistency module	It checks the consistency of data with a specific control rules given in the data set of control rules. For example, PLCs control the water system and give control commands based on the collected data from sensors and actuators.	Each module from all these three modules works separately to identify the attacks. Then, the final result of detecting attacks will be assembling from all the three modules since all the modules are connected by logical statements.
	Pattern recognition module	It contains distinct patterns for hydraulic parameters. There is a pattern for every hydraulic parameter and also there is a pattern for combination of multiple hydraulics parameters. These patterns are developed based on datasets free from cyber-attacks. If the WDS dataset does not follow the developed patterns, the anomalies occur.	
	Hydraulic and system relationships' module	It is developed based on the relationships of WDS components. For instance, the physical information of the water level in the tank and pump station flow can be calculated to derive the information of demand from another component such as mass balance. The calculated values of these components are compared with the collected data of WDS components to detect the attacks.	
[39]	verification of actuator Rules	It ensures that the operations of the pumps and valves follow the right control rule based on the observed water levels in each tank. For example, if the water level in A tank is less than the threshold, the pump should be turned on.	First, the verification of actuator rules checks the operation with their control rules. Then, the statistical method compares the data of sensors and actuators with high and low boundaries. After that, the ANN model trains the data to predict the future observations and discover the anomalies. Finally, the PCA algorithm splits dataset as normal and abnormal.
	Statistical	It monitors the sensors and actuators based on calculating the upper and lower boundaries. If any value is above the upper boundary or below the lower boundary, it will be considered as outliers.	
	ANN model	The dataset will be used to train the ANN model to predict the future observations of tank level data, pressure, and pumping flow rate.	
	PCA	It remaps the multi-dimensional sensor data to new axes called Principal Components (PCs) so that they can separate into two datasets. One data set has a maximum variance of all data and the method classifies the data set as normal data. The other data set has the lowest variance of data and the method classifies the data set as anomalies.	
[40]	Long Short-Term Memory Recurrent Neural Network (LSTM-RNN)	The LSTM algorithm is used to solve the vanishing gradient problem in RNN and increases the ability of model learning. LSTM includes a memory block which consists of an input gate, a forget gate, and an output gate to increase the ability of model to learn data and predict results.	First, LSTM-RNN trains data and predicts the output. Then, Cumulative Sum method compares the predicted value with actual sensor value.
	Cumulative Sum	It is a statistical method used to calculate high and low boundaries.	
[41]	Autoencoder neural networks (AE)	It includes an encoder and a decoder, where the encoder encodes and processes the input data to reduce dimensionality and the decoder reconstructs data from encoded data and obtains a small reconstruction error measured from all data. If the reconstruction error is larger than the small reconstruction error, an anomaly occurs.	N/A
[42]	Rules-based model	It checks the operations and their rules. For example, it will check the water level in each tank with their boundary.	First, the rules-based model checks the operation and checks if the rule is followed correctly. Then, Convolutional variational auto-encoder classifies data as normal or abnormal.
	Convolutional variational auto-encoder	It is one type of deep learning algorithms. It contains encoder and decoder to detect anomalies based on calculating lower and higher values.	
[43]	Simple method	It is mathematical method that calculate the relationship between the physical component and the properties of component.	N/A

TABLE 4. General comparison of detection methods.

# Ref	Target of attacks	Type of method	Results	Advantages	limitation
[13]	Sensors, Actuators, SCADA, PLCs, Communication links between sensors and PLCs, among multiple PLCs, and between PLCs and SCADA.	A simple algorithm	WDS is largely impacted during cyber physical attacks.	The attacks can be used to test detection algorithms.	No good solution to detect attacks.
[30]	Sensors, Actuators, Communication links between actuators and PLCs and among multiple PLCs.	An anomaly behavior detection algorithm	Identifying all the labeled attacks without delay	High Performance	False alarm
[34]	Sensors, Actuators, PLCs, Communication links between sensors and PLCs, between actuators and PLCs, and among multiple PLCs.	An anomaly behavior detection algorithm	Identifying all the labeled attacks	High Performance	The optimization algorithm is not very effective as Data Verification algorithm.
[35]	Sensors and Actuators	An anomaly behavior detection algorithm	Identifying all the labeled attacks	High Performance	The performance decreases when increasing the number of layers of RBM model. The reliability is affected due to lose data during convert real data to binary data in the training RBM model.
[38]	Sensors and Actuators	An anomaly behavior detection algorithm	Identifying all the labeled attacks	High Performance	Large time overhead
[39]	Sensors and Actuators	An anomaly behavior detection algorithm	Identifying all the labeled attacks	High Performance	False alarm
[40]	Sensors, PLCs and Actuators	An anomaly behavior detection algorithm	Identifying all the labeled attacks expect one attack	High Performance	The method is tested with a small dataset training and not entire the dataset of the water system.
[41]	Sensors, Actuators, SCADA, PLCs, Communication links between PLCs and SCADA.	An anomaly behavior detection algorithm	WDS is largely impacted during cyber physical attacks.	It detects all the labeled attacks with define the location of compromised components.	The performance decreases when increasing the number of layers of AE algorithm.
[42]	SCADA Data	An anomaly behavior detection algorithm	It has ability to detect the attacks.	High Performance	The algorithm is applied on partial dataset and not entire the dataset.
[43]	Communication links between PLCs and PLCs, or among multiple PLCs.	A simple detection algorithm.	It detects all the attacks.	It is an effective algorithm	It needs to extend in the future by developing fingerprinting wireless network.

[35], and [38] only use some of the attacks. The paper [30] adopts only four types of attacks and the paper [34] adopts 6 types of attacks. Moreover, both the paper [35] and the paper [38] adopt the same types of attacks that aim to compromise the sensors and actuators. Also, both the paper [13] and the paper [34] adopt the same types of attacks that aim to compromise the sensors, actuators, PLCs, communication links between sensor and PLCs, communication links between actuators and PLCs, and Communication links of multiple PLCs. We observe from Table 1 that only the paper [13] defines SCADA attacks and attacks on communication links between PLCs, which can disable completely the WDS. Table 1 also shows that several papers adopt the same types of attacks that target compromise sensors, actuators, communication links between actuators and PLCs, and communication links of multiple PLCs.

Table 3 compares several papers and their details. The paper [13] focuses on building an attack model by using a simple method without providing a detection solution. However, the attacks model can be used in the papers [30], [34], [35], and [38] that aim to develop anomaly detection algorithms to detect cyber-physical attacks. Specifically, The paper [13] creates nine classes and identifies the features for each attack including the action, the start time, and the end time. The paper [30] adopts 3 layers to detect attacks by using the statistical method, the ANN model, and the PCA method, sequentially. The paper [30] also shows that the statistical method and the ANN model are used to detect attacks from individual sensors while the PCA method detects attacks from multiple sensors at the same time in a short time. The paper [34] adopts Actuators Rules Verification algorithm, Data verification, and an optimization algorithm. The paper [34] also shows that the optimization algorithm

minimizes the detection attack time. We observe from Table 3 that the paper [35] adopts the RBM model and the FFNN model to detect attacks. All of the RBM model, the FFNN model, and the ANN model are artificial neural networks. However, the main difference between the RBM model and the FFNN model is that the RBM model includes only one input layer and one or more hidden layers, while the FFNN model and ANN model include one input layer, one or more hidden layers, and an output layer. The FFNN model is one type of ANN model and both of them can use linear or nonlinear transformation, but the RBM model is always nonlinear transformation. The RBM model must convert real data to binary data which causes loss of information, while the FFNN model can receive binary or real data. The RBM model adopts to learn a large number of features and it has an energy function to reduce the energy. However, it cannot classify the data. The FFNN model can learn features and classify the data, but the accuracy is less than the RBM model. Table 3 also shows that both the PCA method and the RBM model are used to reduce the high dimensional data but the RBM model is better and faster than the PCA method in the dimensionality reduction. The paper [38] adopts a control rule and consistency module, a pattern recognition module, and a hydraulic and system relationships module that works together to identify all attacks.

Table 4 shows a comparison among several papers based on different selection criteria. The authors in [13] propose nine types of attacks and show the importance of developing detection algorithms. As we explained earlier, only the paper [13] does not provide a solution to detect attacks, while others papers develop an algorithm to detect cyber-physical attacks. Table 4 also shows that the paper [30] first adopts a simple statistical method that can easily calculate the results. However, there is a probability to detect false outliers. Then, the paper [30] also adopts the ANN model that is appropriate to learn large numbers of features, but it can suffer from the overfitting problem that reduces the reliability of the model to detect attacks. The overfitting problem occurs when the model is trained with a large number of data and learns noise so that the accuracy of the predicted results reduces. To avoid the overfitting problem in The paper [30], the PCA method is adopted to reduce the number of features to the most important features. However, there is less risk of some information that affects the prediction results. The paper [30] has a false alarm for a long time. The paper [34] shows that all the actuators' rules and data verification are simple algorithms and cannot identifies the complex types of attacks. The actuator rules algorithm can identify all the eight label attacks, while the data verification algorithm identifies six label attacks. The paper [34] targets to extract low dimensional data sensors completely from SCADA and to minimize the computational time data by adopting an optimization algorithm. However, the optimization algorithm cannot extract all the low-dimensional data sensors. The paper [35] adopts the RBM model to increase the accuracy to detect attacks. However, the RBM model must convert real

data sensors to binary data which causes losing information. Moreover, The paper [35] adopts the FFNN model that is suitable to learn a large number of features, but the overfitting problem still risky for the FFNN model. Also, the FFNN model is memoryless and forgets the learning features after many stages in the training data. The authors in [38] adopt three modules to detect attacks, but it has a large time overhead. The paper [34], the paper [35], and the paper [38] can identify all the compromised components, while the paper [30] identifies some of the compromised components.

In summary, integrating physical water infrastructure with cyber systems exposes the WDS to cyber-physical attacks. As we indicated above, there are some cyber-physical attacks and detection algorithms available. However, many detection algorithms are still not optimal for identifying all the attacks. Furthermore, there are inadequate automated monitoring and reporting that allow appropriate responses in a short time to detect or mitigate attacks.

V. FUTURE RESEARCH DIRECTIONS

In the above sections, we observe the following conclusions. WDS is a critical infrastructure and we cannot continue normal life without water. However, Cyber-physical attacks are big challenges and can cause severe impacts on the security level of WDS. We summarize many cyber-physical attacks and their impacts on the WDS and many detection methods. We observe that the number of attacks is still limited and we believe that in the future, more attacks will emerge. Furthermore, the existing detection methods are not optimal and most of them could not detect all the attacks. In the future, there are many research directions that people (including us) can work on them.

- First, designing accurate, better detection methods is important. Most of the current methods adopt machine learning or artificial neural network algorithms. However, high dimensional data and ensemble training data set have some drawbacks of these approaches. Moreover, ensemble training data set can affect the accuracy of the algorithms to identify the attacks especially in the detection methods that adopt machine learning or artificial neural networks. Enhancing the data training process by reducing the false-negative rate is still a big challenge. Also, hybrid methods using the advantages of different methods deserve better studies. Furthermore, since all the surveyed papers cannot detect all the attacks, designing a comprehensive detection method to detect all attacks is deserved more studies.
- Second, we can improve the security of network and communication to increase the reliability of exchanging data among different components of WDS. There are heterogeneous communication networks to exchange data such as Home Area Network (HAN), Wide Area Network (WAN), and wireless communication (Wi-Fi), etc. However, the diversity of network communication

types can cause a big challenge to the security of WDS. There are many problems such as heterogeneous network authentication, access control, privacy protection, key management, information storage, and security devices that should be studied in the future. Furthermore, Modern WDS not only exchange data over networks but also store and process these data in central systems such as SCADA systems. We observe that many attacks target SCADA systems to control completely the water system [15]. Proper cryptography tools are needed to be deployed in proper locations in the WSD. Where and how to deploy these tools in the WDS need careful studies to prevent attacks.

- Third, Internet of Things (IoT) with other smart devices empower the WDS by allowing transparency to the processes in the water supply chain, ensuring real-time monitoring, and automatic processing. However, the WDS becomes more vulnerable to attackers due to the limitation of resources. Both security and privacy are still major problems especially in IoT devices due to their attributes such as the low battery, small memory, small capability, lack of regulations, etc. The privacy violation reduces the level of acceptance of the water system generally. The major aim of privacy violation is to expose sensitive information to unreliable and undependable parties after achieving the feat of getting access to the data by hacking it through the unguided portion of the IoT system. For example, user information on water consumption can be exposed when a smart meter reads the water consumption by the user several times. Also, the SCADA system can be kept in advanced technology like cloud and fog to process the consumption data and store that personal information, financial information, and produce the bill. However, security and privacy are also still challenging and there is the probability to expose and breach the dataset of water consumers. Furthermore, some of the smart water devices have their geographical location accessible. Thus, an attacker may effortlessly get unrestricted authorization to the device to steal the information. There isn't a universal security standard for IoT technologies. An appropriate level cryptographic algorithm should be designed for the WDS that can be suitable with the limitation of new smart technologies such as low power, small storage, small size, etc.

- Fourth, most of the attention has been paid to quantify the impact of attacks on the WDS. However, a little resilience study is achieved to build WDS with more resilience and recover faults and attacks. Faults can occur due to failure or faults of main components such as sensors, PLCs, SCADA systems, etc. The failure might trigger many issues such as pipe breaks, tank failure, pump outage, and valve locking. The current WDS deploys redundant software or hardware to recover faults, but this is very costly. WDS should be designed to be more resilient and fault-tolerant when faults

and attacks happen. Furthermore, WDS is a real-time system and should be operating continuously even after a fault occurs. It is important to develop fault/attack tolerance techniques that can increase resilience, isolate the faulty/attacked components, and recovery from failures/attacks to minimize the bad effects. Deep learning algorithms can be adopted as fault-tolerant solutions and more studies should be conducted to increase the ability to determine exact locations of faults/attacks and to provide mitigation and recovery.

- Finally, current attacks on water systems are limited and are mostly similar among different papers. Designing novel attacks on water systems are also needed to test the security and robustness of the systems. Collaborated attacks, multiple-factor attacks, and anti-detection attacks can be designed on water systems as a future research direction.

In summary, the WDS can be enhanced by prevention, detection, mitigation, and recovery mechanisms. Both secure software and hardware need to be considered. Cryptographic algorithms, security network/ communication protocols, cyber-physical system designs, device-level security (such as IoT security), etc., should be considered. There is still very limited research on security for WDS in the literature as indicated in this paper. There are a lot of research topics that can be carried out in the future.

VI. CONCLUSION

A Water Distribution System (WDS) is a critical infrastructure for providing high-quality water services by transferring clean water to consumers and recycling the dirty water back. It operates based on the main components, such as pipes, tanks, sensors, PLCs, SCADA systems, etc. However, WDS is prone to cyber-physical attacks that impact the basic operation and disrupt the service. Therefore, it is necessary to develop and implement solutions to increase the security of the WDS by preventing, detecting, mitigating, and recovering cyber-physical attacks. In this paper, we provide a comprehensive survey that shows the impacts of cyber-physical attacks on the WDS and we survey the common detection methods that mitigate the negative impacts of cyber-physical attacks. We analyze them with details, especially on methods, evaluation results, and limitations. We also compare them based on different evaluation criteria and provide future research directions. We realize that there is a shortage in the security research of WDS and we hope that our paper can trigger more research to increase the security of WDS against these attacks and threats.

REFERENCES

[1] J. R. Gil-Garcia, T. A. Pardo, and T. Nam, "What makes a city smart? Identifying core components and proposing an integrative and comprehensive conceptualization," *Inf. Polity*, vol. 20, no. 1, pp. 61–87, Jul. 2015.

[2] P. Lynggaard, "Using neural networks to reduce sensor cluster interferences and power consumption in smart cities," *Int. J. Sens. Netw.*, vol. 32, no. 1, pp. 25–33, 2020.

[3] C. Harrison, B. Eckman, R. Hamilton, P. Hartswick, J. Kalagnanam, J. Paraszczak, and P. Williams, "Foundations for smarter cities," *IBM J. Res. Develop.*, vol. 54, no. 4, pp. 1–16, Jul./Aug. 2010.

[4] S. P. Mohanty, U. Choppali, and E. Kougianos, "Everything you wanted to know about smart cities: The Internet of Things is the backbone," *IEEE Consum. Electron. Mag.*, vol. 5, no. 3, pp. 60–70, Jul. 2016.

[5] O. Shafaiy and N. Farzaneh, "Smart speed advisory system for drivers with priority assignment for smart city," *Int. J. Sensor Netw.*, vol. 36, no. 2, p. 68, 2021.

[6] UNESCO, "The united nations world water development report 2019," UNESCO, Paris, France, United Nations World Water Develop. Rep. 3, 2019, p. 36.

[7] M. Mutchek and E. Williams, "Moving towards sustainable and resilient smart water grids," *Challenges*, vol. 5, no. 1, pp. 123–137, Mar. 2014.

[8] A. Gonzalez-Vidal, J. Cuenca-Jara, and A. F. Skarmeta, "IoT for water management: Towards intelligent anomaly detection," in *Proc. IEEE 5th World Forum Internet Things (WF-IoT)*, Apr. 2019, pp. 858–863.

[9] V. Hopman, P. Kruiver, A. Koelewijn, and T. Peters, "How to create a smart levee," in *Proc. 8th Int. Symp. Field Meas. GeoMechanics*, 2010, pp. 12–16.

[10] J. Li, X. Yang, and R. Sitzenfrei, "Rethinking the framework of smart water system: A review," *Water*, vol. 12, no. 2, p. 412, Feb. 2020.

[11] S. K. S. G. Priya and T. Revathi, "Design of smart sensors for real time drinking water quality monitoring and contamination detection in water distributed mains," *Int. J. Eng. Technol.*, vol. 7, no. 1, pp. 47–51, 2018.

[12] H. N. Saha, S. Auddy, A. Chatterjee, S. Pal, S. Sarkar, R. Singh, A. K. Singh, P. Sharan, S. Banerjee, R. Sarkar, and A. Maity, "IoT solutions for smart cities," in *Proc. 8th Annu. Ind. Autom. Electromech. Eng. Conf. (IEMECON)*, Aug. 2017, pp. 74–80.

[13] R. Taormina, S. Galelli, N. O. Tippenhauer, E. Salomons, and A. Ostfeld, "Characterizing cyber-physical attacks on water distribution systems," *J. Water Resour. Planning Manage.*, vol. 143, no. 5, May 2017, Art. no. 04017009.

[14] B. Sun, F. Ahmed, F. Sun, Q. Qian, and Y. Xiao, "Water quality monitoring using STORM 3 data loggers and a wireless sensor network," *Int. J. Sens. Netw.*, vol. 20, no. 1, pp. 26–36, 2016.

[15] J. Gao, J. Liu, B. Rajan, R. Nori, B. Fu, Y. Xiao, W. Liang, and C. L. P. Chen, "SCADA communication and security issues," *Secur. Commun. Netw.*, vol. 7, no. 1, pp. 175–194, Jan. 2014.

[16] M. Abrams and J. Weiss, *Malicious Control System Cyber Security Attack Case Study-Maroochy Water Services, Australia*. McLean, VA, USA: The MITRE Corporation, 2008.

[17] D. Nikolopoulos, G. Moraitis, D. Bouziotas, A. Lykou, G. Karavokiros, and C. Makropoulos, "Cyber-physical stress-testing platform for water distribution networks," *J. Environ. Eng.*, vol. 146, no. 7, 2020, Art. no. 04020061.

[18] A. Vera, J. Lynch, and C. Carrega, "Someone tried to poison a Florida city by hacking into the water treatment system, sheriff says," Cable News Netw., Atlanta, GA, USA, Feb. 2021. Accessed: Mar. 10, 2021. [Online]. Available: https://www.cnn.com/2021/02/08/us/oldsmar-florida-hack-water-poison/index.html

[19] E. Bou-Harb, "Passive inference of attacks on SCADA communication protocols," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[20] A. Ercumen, J. S. Gruber, and J. M. Colford, Jr, "Water distribution system deficiencies and gastrointestinal illness: A systematic review and meta-analysis," *Environ. Health Perspect.*, vol. 122, no. 7, pp. 651–660, 2014.

[21] R. Taormina, S. Galelli, N. O. Tippenhauer, E. Salomons, A. Ostfeld, D. G. Eliades, and M. Aghashahi, "Battle of the attack detection algorithms: Disclosing cyber attacks on water distribution networks," *J. Water Resour. Planning Manage.*, vol. 144, no. 8, 2018, Art. no. 04018048.

[22] S. Hajebi, H. Song, S. Barrett, A. Clarke, and S. Clarke, "Towards a reference model for water smart grid," *Int. J. Adv. Eng. Sci. Technol.*, vol. 2, no. 3, pp. 310–317, 2013.

[23] D. P. Loucks and E. van Beek, "Urban water systems," in *Water Resource Systems Planning and Management*. 2017, pp. 527–565.

[24] S. Diaz, A. Molano, C. Erazo, and J. C. Monroy, "WQMS: Water quality monitoring station for IoT," *Int. J. Sens. Netw.*, vol. 35, no. 2, pp. 79–87, 2021.

[25] T. K. Chai, "Security for smart cities," *IET Smart Cities*, vol. 2, pp. 95–104, Aug. 2020.

[26] C. M. Ahmed, V. R. Palleti, and A. P. Mathur, "WADI: A water distribution testbed for research in the design of secure cyber physical systems," in *Proc. 3rd Int. Workshop Cyber-Phys. Syst. Smart Water Netw.*, 2017, pp. 25–28.

[27] D. I. Urbina, J. A. Giraldo, N. O. Tippenhauer, and A. A. Cárdenas, "Attacking fieldbus communications in ICS: Applications to the SWaT testbed," in *Proc. SG-CRC*, 2016, pp. 75–89.

[28] A. Hassanzadeh, A. Rasekh, S. Galelli, M. Aghashahi, R. Taormina, A. Ostfeld, and M. K. Banks, "A review of cybersecurity incidents in the water sector," *J. Environ. Eng.*, vol. 146, no. 5, 2020, Art. no. 03120003.

[29] G. Bernieri, E. E. Miciolino, F. Pascucci, and R. Setola, "Monitoring system reaction in cyber-physical testbed under cyber-attacks," *Comput. Elect. Eng.*, vol. 59, pp. 86–98, Apr. 2017.

[30] A. A. Abokifa, K. Haddad, C. S. Lo, and P. Biswas, "Detection of cyber physical attacks on water distribution systems via principal component analysis and artificial neural networks," in *Proc. World Environ. Water Resour. Congr.*, 2017, pp. 676–691.

[31] Z. Fountas, "Spiking neural networks for human-like avatar control in a simulated environment," Student's thesis, Comput. Sci., Imperial College London, London, U.K., Sep. 2011.

[32] A. Ali. *Restricted Boltzmann Machine (RBM) With Practical Implementation*. Accessed: Oct. 10, 2020. [Online]. Available: https://medium.com/machine-learning-researcher/boltzmann-machine-c2ce76d94da5

[33] J. Fang, B. Li, and M. Gao, "Collaborative filtering recommendation algorithm based on deep neural network fusion," *Int. J. Sens. Netw.*, vol. 34, no. 2, pp. 71–80, 2020.

[34] M. Giacomoni, N. Gatsis, and A. Taha, "Identification of cyber attacks on water distribution systems by unveiling low-dimensionality in the sensory data," in *Proc. World Environ. Water Resour. Congr.*, 2017, pp. 660–675.

[35] A. Elsaeidy, K. S. Munasinghe, D. Sharma, and A. Jamalipour, "A machine learning approach for intrusion detection in smart cities," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2019, pp. 1–5.

[36] *Feed Forward Neural Network Learned by a GA*. Accessed: Oct. 10, 2020. [Online]. Available: https://mangans84.wordpress.com/tag/feedforward-neural-network/

[37] T. Mahjabin, Y. Xiao, G. Sun, and W. Jiang, "A survey of distributed denial-of-service attack, prevention, and mitigation techniques," *Int. J. Distrib. Sensor Netw.*, vol. 13, no. 12, pp. 1–33, Dec. 2017, doi: 10.1177/1550147717741463.

[38] M. F. K. Pasha, B. Kc, and S. L. Somasundaram, "An approach to detect the cyber-physical attack on water distribution system," in *Proc. World Environ. Water Resour. Congr.*, 2017, pp. 703–711.

[39] A. A. Abokifa, K. Haddad, C. Lo, and P. Biswas, "Real-time identification of cyber-physical attacks on water distribution systems via machine learning–based anomaly detection techniques," *J. Water Resour. Planning Manage.*, vol. 145, no. 1, Jan. 2019, Art. no. 04018089.

[40] J. Goh, S. Adepu, M. Tan, and Z. S. Lee, "Anomaly detection in cyber physical systems using recurrent neural networks," in *Proc. IEEE 18th Int. Symp. High Assurance Syst. Eng. (HASE)*, Jan. 2017, pp. 140–145.

[41] R. Taormina and S. Galelli, "Deep-learning approach to the detection and localization of cyber-physical attacks on water distribution systems," *J. Water Resour. Planning Manage.*, vol. 144, no. 10, Oct. 2018, Art. no. 04018065.

[42] S. E. Chandy, A. Rasekh, Z. A. Barker, B. Campbell, and M. E. Shafiee, "Detection of cyber-attacks to water systems through machine-learning-based anomaly detection in SCADA data," in *Proc. World Environ. Water Resour. Congr.*, 2017, pp. 611–616.

[43] S. Adepu, J. Prakash, and A. Mathur, "WaterJam: An experimental case study of jamming attacks on a water treatment system," in *Proc. IEEE Int. Conf. Softw. Qual., Rel. Secur. Companion (QRS-C)*, Jul. 2017, pp. 341–347.

[44] N. Nicolaou, D. G. Eliades, C. Panayiotou, and M. M. Polycarpou, "Reducing vulnerability to cyber-physical attacks in water distribution networks," in *Proc. Int. Workshop Cyber-physical Syst. Smart Water Netw. (CySWater)*, Apr. 2018, pp. 16–19.

[45] M. Housh and Z. Ohar, "Model-based approach for cyber-physical attack detection in water distribution systems," *Water Res.*, vol. 139, pp. 132–143, Aug. 2018.

[46] M. Housh and Z. Ohar, "Integrating physically based simulators with event detection systems: Multi-site detection approach," *Water Res.*, vol. 110, pp. 180–191, Mar. 2017.

[47] Y. Chen, C. M. Poskitt, J. Sun, S. Adepu, and F. Zhang, "Learning-guided network fuzzing for testing cyber-physical system defences," in *Proc. 34th IEEE/ACM Int. Conf. Automated Softw. Eng. (ASE)*, Nov. 2019, pp. 962–973.

[48] A. Elsaeidy, K. S. Munasinghe, D. Sharma, and A. Jamalipour, "Intrusion detection in smart cities using restricted Boltzmann machines," *J. Netw. Comput. Appl.*, vol. 135, pp. 76–83, Jun. 2019.
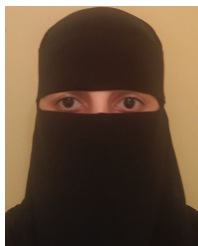
HAJAR HAMEED ADDEEN received the bachelor's degree in computer science from Taif University, Taif, Saudi Arabia, and the master's degree from South Dakota State University, Brookings, SD, USA, in 2019. She is currently pursuing the Ph.D. degree with The University of Alabama. Her research interests include computer security, network security, and the Internet of Things. She is a member of Upsilon Pi Epsilon (UPE) for computing and information disciplines.

YANG XIAO (Fellow, IEEE) received the B.S. and M.S. degrees in computational mathematics from Jilin University, Changchun, China, and the M.S. and Ph.D. degrees in computer science and engineering from Wright State University, Dayton, OH, USA. He is currently a Full Professor with the Department of Computer Science, The University of Alabama, Tuscaloosa, AL, USA. His current research interests include cyber-physical systems, the Internet of Things, security, wireless networks, smart grid, and telemedicine. He has published over 300 SCI-indexed journal articles (including over 50 IEEE/ACM transactions articles) and 250 EI indexed refereed conference papers related to these research areas. He was a Voting Member of the IEEE 802.11 Working Group, from 2001 to 2004, involving the IEEE 802.11 (WiFi) standardization work. He is an IET Fellow. He currently serves as the Editor-in-Chief for *Cyber-Physical Systems* journal. He has served as an Editorial Board or an Associate Editor for 20 international journals, including the IEEE TRANSACTIONS ON CYBERNETICS, since 2020, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS (2014–2015), IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (2007–2009), and IEEE COMMUNICATIONS SURVEY AND TUTORIALS (2007–2014). He has served as a Guest Editor for over 20 times in different international journals, including the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE NETWORK, IEEE WIRELESS COMMUNICATIONS, and ACM/Springer *Mobile Networks and Applications* (MONET).
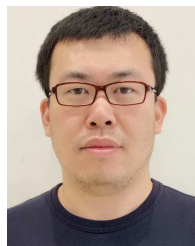
JIACHENG LI (Student Member, IEEE) received the B.S. degree in computer science from California State University, Fresno, CA, USA. He is currently pursuing the Ph.D. degree with the Department of Computer Science, The University of Alabama, Tuscaloosa, AL, USA. His research interests include accountability for the Internet of Things (IoT), cyber security, and blockchain.

MOHSEN GUIZANI (Fellow, IEEE) received the B.S. (Hons.), M.S., and Ph.D. degrees in electrical and computer engineering from Syracuse University, Syracuse, NY, USA. He is currently a Professor with the Computer Science and Engineering Department, Qatar University, Qatar. Previously, he worked in different institutions, such as the University of Idaho, Western Michigan University, the University of West Florida, the University of Missouri-Kansas City, the University of Colorado Boulder, and Syracuse University. He is the author of nine books and more than 800 publications. His research interests include wireless communications and mobile computing, applied machine learning, cloud computing, and security and its application to healthcare systems. He was elevated to an IEEE Fellow, in 2009. He was listed as a Clarivate Analytics Highly Cited Researcher in Computer Science, in 2019 and 2020. He has won several research awards, including the 2015 IEEE Communications Society Best Survey Paper Award and four best paper awards from ICC and GLOBECOM conferences. He was also a recipient of the 2017 IEEE Communications Society Wireless Technical Committee (WTC) Recognition Award, the 2018 *Ad Hoc* Technical Committee Recognition Award, and the 2019 IEEE Communications and Information Security Technical Recognition (CISTC) Award. He served as the Editor-in-Chief for *IEEE Network* and currently serves on the Editorial Board of many IEEE journals/transactions. He was the Chair of the IEEE Communications Society Wireless Technical Committee and the TAOS Technical Committee. He served as the IEEE Computer Society Distinguished Speaker and is currently the IEEE ComSoc Distinguished Lecturer.

• • •