doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14799

PATHOINVEST: PATHOGEN CONTAMINATION INVESTIGATIONS DURING EMERGENCIES

Sotirios Paraskevopoulos[1], Stelios Vrachimis[2], Marios Kyriakou[3], Pavlos Pavlou[4], Dimitris Kouzapas[5], George Milis[6], Patrick Smeets[7], Demetrios G. Eliades[8], Gertjan Medema[9], Marios Polycarpou[10] and Christos Panayiotou[11]

[1,2,3,4,5,8,10,11]KIOS Research and Innovation Center of Excellence, University of Cyprus, Nicosia, Cyprus
[10,11]Electrical and Computer Engineering Department, University of Cyprus, Nicosia, Cyprus
[7,9]KWR Watercycle Research Institute, Nieuwegein, The Netherlands
[6] Department of Water Management, Delft University of Technology, Delft, The Netherlands
PHOEBE Research and Innovations Ltd, Cyprus

[1] *Sotirios.Paraskevopoulos@kwrwater.nl*, [2] *vrachimis.stelios@ucy.ac.cy*, [3] *kiriakou.marios@ucy.ac.cy*, [4] *pavlou.v.pavlos@ucy.ac.cy*, [5] *kouzapas.dimitrios@ucy.ac.cy*, [6] *george.milis@phoebeinnovations.com*, [7] *patrick.smeets@kwrwater.nl*, [8] *eldemet@ucy.ac.cy*, [9] *gertjan.medema@kwrwater.nl*, [10] *mpolycar@ucy.ac.cy*, [11] *christosp@ucy.ac.cy.*

Abstract

Emergencies and disasters (such as earthquakes and floods), may contaminate drinking water systems with pathogens, that can affect the health of both First Responders and Citizens. As part of the Horizon 2020 "Pathogen Contamination Emergency Response Technologies" (PathoCERT) project, we are developing a Digital Twin tool (PathoINVEST) to assist First Responders and Water Authorities in investigating and responding efficiently to drinking water contamination events. In this paper, we present preliminary work on PathoINVEST, its architecture, and how it operates with the PathoCERT ecosystem of technologies. Moreover, using an illustrative case study, we demonstrate how PathoINVEST will process data and produce useful insights for the First Responders during a realistic contamination event. This work demonstrates how different research results can be integrated into a holistic water contamination emergency management system, in accordance with the needs of First Responders who need to make decisions within a limited time frame and to reduce the impact of a contamination event.

Keywords
Water contamination, first responders, emergency response, digital twins.

1 INTRODUCTION

During disasters and emergencies, water systems can be unexpectedly exposed to pathogens [1, 2]. For example, during an earthquake, the drinking and sewerage system of a city may be affected by pipe breaks, which may cause infiltration of sewage water into the drinking water network. Floods after intense rainfall may carry away toxic substances, whereas, in the case of technological accidents, or even bioterrorism attacks, unknown pathogens may be injected into the drinking water supply.

During these emergencies, citizens, as well as First Responders operating in the area, may become exposed to contaminated water, through skin contact, ingestion, or inhalation. This can pose a significant risk of illness, disease, or even death. In these situations, water may be contaminated with pathogens such as Cryptosporidium, E. coli O157:H7, Norovirus, and Vibrio cholerae [3]. Some recent examples of these emergency events include an earthquake causing infiltration of dirt in the Drinking Water Network (DWN) (Larisa, Greece, March 2021), flash-flood water overflowing a wastewater treatment plant which went through the city of Merrit (British

Columbia, Canada, November 2021), and hospital sewage infiltrating a water distribution network (Prague, May 2015).

To effectively manage these situations, and to reduce the loss of human lives, First Responders and Water Authorities need to be equipped with the appropriate emergency response technologies. This is also the goal of the project "Pathogen Contamination Emergency Response Technologies" (PathoCERT), which is funded by the European Union under the Horizon 2020 programme. PathoCERT aims to develop new technologies for fast alerting and detection of pathogens in surface and drinking water, using smart portable sensors, as well as develop new technologies for improving situational awareness using drones, satellites, social media, and smart cameras. A new intelligent IoT gateway and an open-standards (FIWARE-based) platform will be designed to support the collection and analysis of these heterogeneous data. The platform is connected with tools for threat and risk assessment, as well as for investigating contamination events and proposing the most appropriate mitigation actions. Eventually, these tools will increase the First Responder and Water Operator capabilities and reduce the exposure to pathogens during an emergency, which can have a significant impact on human health.

This paper presents an overview of related research on the topic of water contamination emergency response management and introduces PathoINVEST, a Digital Twin Platform for investigating urban drinking water contamination events. PathoINVEST is a collection of tools that incorporates sensor telemetry and hydraulic models to create an up-to-date state-estimation of the network model. Additionally, it uses a set of modeling tools for forecasting contaminant evolution during emergencies, health risk assessment, mobile/portable sensor deployment, and evaluation of mitigation measures to reduce exposure and health risk.

The main contributions of this work are a) the presentation of the PathoINVEST reference architecture, and b) the release of a complete contamination emergency management case study using the L-Town benchmark network.

This paper is structured as follows: In Section 2, a literature review on pathogen contamination emergency response technologies is provided highlighting the key elements that constitute an emergency response (from preparedness and contamination modeling to assessing the risk and applying mitigation measures). In Section 3, the architecture of the PathoINVEST digital twin system and the PathoCERT ecosystem is presented, whereas Section 4 presents an illustrative case study using the L-Town benchmark and example implementations, of how a possible contamination event could be managed using the proposed tool.

2 BACKGROUND ON PATHOGEN CONTAMINATION EMERGENCY RESPONSE TECHNOLOGIES

To deal with emergencies, such as floods, earthquakes, accidents, or even attacks, First Responders and Water Authorities need to coordinate to effectively manage the situation. To achieve this, procedures and technologies need to be in place to capture the different phases of the event. These phases include: 1) monitoring and detecting contamination events, 2) assessing the threat and risk, 3) identifying the source, and 4) mitigating the contamination event. For each phase, significant research has been conducted during the previous years, and this section summarizes some important works.

2.1 Monitoring and detecting contamination events

Due to their high costs, a key challenge is to decide what type of sensors to install, how many, and in which locations in the network, considering all the uncertainties. Another related challenge is where and when to conduct manual sampling for lab analysis. These problems have been studied extensively, and are typically expressed as optimization problems, sometimes with multiple objectives which may include impact risk, detection time, and coverage [4, 5, 6, 7].

After the placement of the sensors, online monitoring tools can be employed to continuously analyze the readings and determine changes in water quality, which could be attributed to contamination events. In the literature, both model-based and data-driven approaches have been proposed, considering one or more water quality parameters [8, 9]. In addition to sensors monitoring physical parameters within water systems, researchers in social networks study how citizens can act as sensors. They focus on what information can be extracted from their networks to improve the ability to determine the extent of a contamination event and to allocate the appropriate response mechanism [10, 11, 12].

2.2 Threat and Risk Assessment

Water utilities are increasingly required to establish Water Security Plans (WSP). That includes preparations to manage emergencies that threaten their system. According to the World Health Organization (WHO) [13], risk assessment is an integral part of developing and implementing a WSP. Many factors can cause adverse effects to this critical infrastructure through the introduction of microbiological, chemical, or radiological hazards. Therefore, risk assessment is considered imperative to facilitate the evaluation of health risks associated with contamination of water supply and to assist responsible authorities in controlling and mitigating an event. One form of risk assessment that has received attention over the last two decades and has been embedded in the WHO water-related guidelines [14] is the Quantitative Microbial Risk Assessment (QMRA). QMRA is a mathematical framework for evaluating infectious risks by combining scientific knowledge about pathogens (fate, transport, route of exposure, and health effects of human pathogens) with the effect of physical/mechanical barriers and mitigation actions [13]. There are 4 steps associated with QMRA, summarized in Table 1.

Table 1. Steps of QMRA (adapted from WHO [13]).

Step	***Description***
1. Problem formulation	Highlight the reference pathogens, exposure pathways, contamination events, and health outcomes of interest
2. Exposure assessment	Measure the dose and frequency of contaminant to which people are exposed (via the identified exposure pathway)
3. Health effects assessment	Identify a dose-response relationship (linking dose to the probability of infection/illness) for the reference pathogen.
4. Risk characterization	Combine the information from the exposure and health effects assessment to generate a quantitative measure of risk

Scientists are researching the combination of QMRA with modeling tools to assess the risk of a drinking water contamination event for the population. The research ranges from main breaks and wastewater intrusion in the network due to pressure transients (unintentional contamination) to hypothetical intentional wastewater injection at key locations. The work from [15] assessed the risk due to wastewater intrusion after negative pressure transients. [16, 17] combined modeling with QMRA using stochastic water demands to account for contamination events after main repairs in the DWN. The authors concluded that the initial contaminant concentration determines highly the exposure, while the infection risk is determined by the most infectious pathogen dose-response. Finally, other works assessed the effects and exposure scenarios after deliberate microbiological contamination in a DWN [18, 19]. Specifically, the work in [18] used a variation of QMRA by modeling the number of affected consumers and highlighting the most critical areas (in terms of exposure) within a network, while [19] investigated the effects of duration, concentration, exposure pathway, and pathogen infectivity on exposure and infection risk.

2.3 Contamination source identification

During chemical or microbial contamination in a DWN, the ability of a water utility to have an early indication of the potential source of contamination is of utmost importance since it can facilitate mitigation measures to stop the spreading and isolate the contaminant. The *Contamination Source Identification* (CSI) problem is mainly considered a deterministic inverse problem, where, using hydraulic calculations, water parcels are backtracked to reach the source of contamination [20]. It is considered a challenging problem due to the computational burden associated with hydraulic calculations, hydraulic uncertainties, and the non-uniqueness of the solutions in identifying the source [20].

A first attempt to solve the CSI was in [21], which introduced the particle backtracking approach. The authors used a particle backtracking algorithm based on a Lagrangian model where the contaminants were considered as particles that run in reverse time from the detection node to the source of contamination.

An alternative response action to tackle a contamination event in the DWN is the option of expanded sampling. Water utilities can focus on examining water quality at specific locations within a DWN (after initial detection of contamination), to help evaluate the contamination impact and identify the potential source area. [22] proposed a computational approach based on decision trees to select a sequence of, as few as possible, nodes for expanded sampling, during which the contamination impact is evaluated, and the source of contamination is isolated/identified.

2.4 Mitigation Measures

After the source identification, a water utility wants to minimize the impact of confirmed contamination by using mitigation measures such as network operational interventions or Optimal Booster Chlorination Placement (OBCP).

Regarding network operational interventions, the most common is valve manipulation. For example, [23] proposed an active contamination detection system by manipulating valves to drive flows to designated nodes within a DWN, thus enabling sensors to monitor water quality. The objective function of this study was to minimize the impact on the population by detecting the contaminant as fast as possible.

Regarding OBCP, [24] presented an optimization model to tackle a (un)intentional contamination event in the DWN. The authors included the reactions of chlorine with unknown contaminants, their fate, and transport, as well as the time delay between detection and application of booster chlorination. The objective was to minimize the impact on the population specified as the number of people who ingest contaminated water above a specified mass threshold.

2.5 Modeling water quality and contaminants in drinking water networks

In most DWN, disinfection is performed to provide (microbiologically) safe drinking water and prevent water quality deterioration. For some of the methodologies described previously, it is assumed that a mathematical/computational model describes the reaction dynamics of the disinfectant agent with contaminants in the water. Both chlorine reactions and pathogen inactivation kinetics can be modeled using *EPANET-MSX*, an open-source multi-species simulator used in conjunction with the *EPANET* hydraulic simulator. EPANET-MSX considers multiple interacting species and enables the modeling of fate, transport, and reaction dynamics of biological species [25].

Various researchers have investigated the interactions of chlorine with microbial contaminants in a DWN. In [18], the authors assessed the vulnerability of a network under deliberate contamination attacks using a parallel first-order model to describe chlorine decay (coupled with a bacterial regrowth model). Moreover, in [26], the authors also modeled parallel first-order (fast and slow) chlorine reactions with microbial contaminants in a DWN. In [27], the researchers modeled chlorine reactions with pathogens (wastewater intrusion) incorporating water quality parameters (pH, temperature), while using *Giardia* and *Escherichia coli* 0157: H7 as the intrusion pathogens. The Chick-Watson equation and first-order kinetics were used to describe the inactivation of the microbial contaminants and chlorine decay respectively.

3 SOFTWARE ARCHITECTURE

The PathoCERT integrated solution is composed of different technological modules (such as sensors) as well as information systems, to assist first responders and other relevant stakeholders in communicating information, sharing knowledge, and effectively managing contamination threats and events. In PathoCERT, all software modules are connected to a FIWARE-enabled backend, the PathoWARE. The vision is for this platform to be installed beforehand and activated when an emergency occurs, following Standard Operating Procedures. Even though the details may be different in the various countries, as a general framework, we will assume that during emergencies, a Command-and-Control Center is set up and there, and a monitoring area with multiple screens is set up and operated by experts. The Center is managed by the Incident Commander, who has the role of coordinating the activities of the different First Responders, as well as the experts and the other relevant stakeholders. Other First Responders and Utility personnel may operate on the field, whereas other experts may provide specialized knowledge and support. Figure 1 illustrates the proposed architecture for the integrated PathoCERT modules and technologies, and in more detail the PathoINVEST module. In the following paragraphs, the different modules are be explained in more depth:

PathoWARE is the core of the PathoCERT Platform, based on FIWARE. It is a cloud-based solution, responsible for collecting data from heterogeneous data sources, harmonizing and processing the data to generate useful information, as well as serving data and results to the other modules of the PathoCERT platform. Data interoperability is achieved using a common PathoCERT ontology. PathoWARE will operate as a service facilitating the integration of different modules from the PathoCERT ecosystem, including the IoT gateways, situation awareness technologies for processing data from social media, as well as interfacing with wearables, mobile apps, GIS, and Decision Support tools for threat and risk assessment. In addition, PathoWARE can provide access to DWN models (provided by a water utility, that can be extracted from the GIS using middleware software).

PathoINVEST is a digital twin of the DWN which implements functionalities that support decision-making during contamination emergencies (Figure 1). These functionalities will be supported by software tools that are integrated into PathoINVEST, including a) state estimation (for estimating the hydraulic states based on the available flow and pressure measurements), b) demand forecasting (to estimate the future hydraulic dynamics), c) simulation tools for multiple species and reactions, d) optimization tools for sampling and sensor placement, e) tools to estimate the health impacts using epidemiological data and population statistics, and f) models that simulate mitigation measures (valves closing, network flushing and booster chlorination). The User Interface of PathoINVEST allows the operator (typically a modeling expert), to receive requests from the Incident Commander, who is responsible for managing the emergency. The PathoINVEST operator will manually set up the software to produce the requested outputs, and decide which information (maps, animations, figures, etc.) to communicate back to the Command-and-Control Center. PathoINVEST is being developed as a QGIS plugin, and the analytics modules will be implemented on Python and MATLAB.

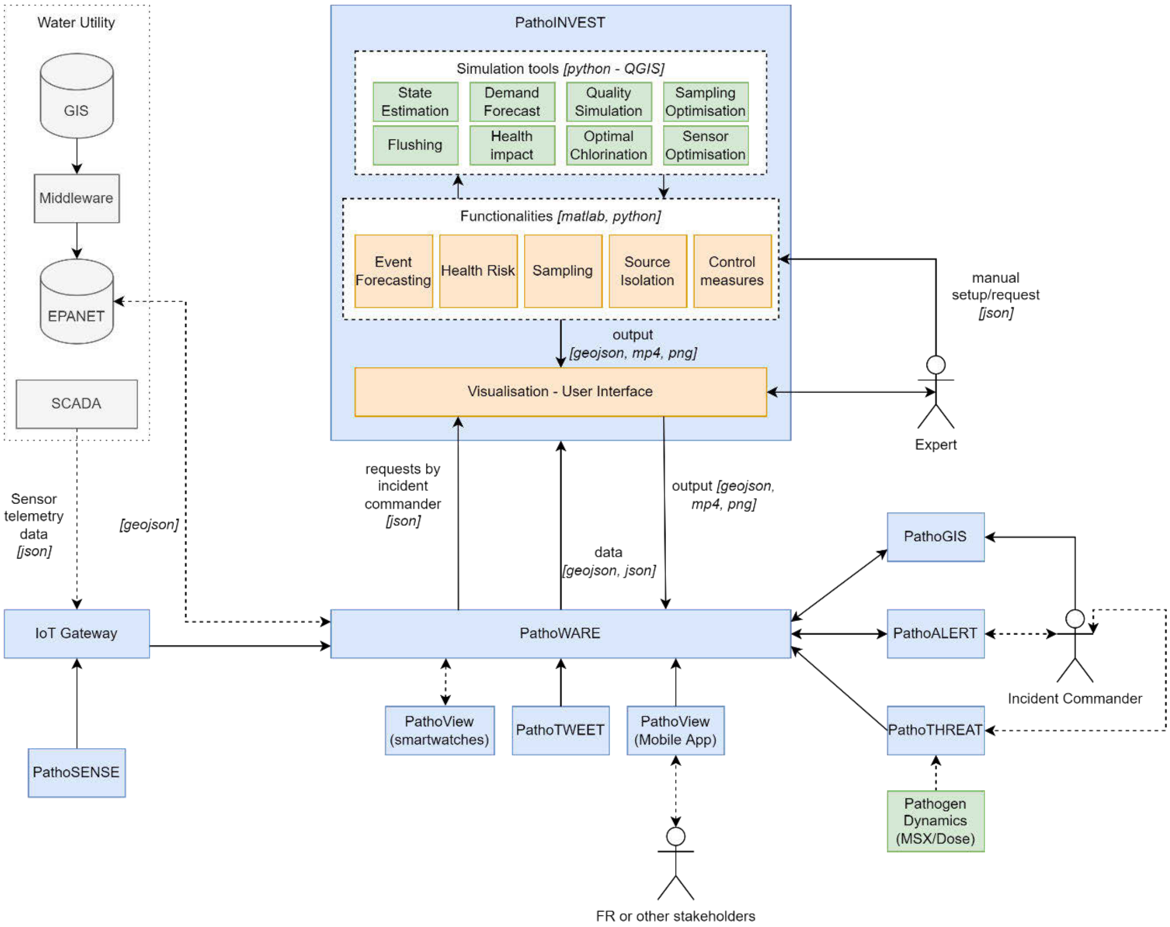

Figure 1. The architecture of PathoINVEST within the PathoCERT ecosystem.

PathoSENSE is a collection of PathoCERT-compatible water quality sensors, which can be used for alerting and detecting pathogens within the DWN, in less than 1 hour. In addition to in-line sensors, portable sensors (**PathoTeSTICK**) which can be carried by First Responders and connected to their mobile phones, are also being developed, to detect specific pathogens within a few minutes [28]. These sensors can connect wirelessly to a specialized ***PathoSENSE IoT Gateway*** (via Wi-Fi, Bluetooth, or LoRaWAN), which manages the metadata of these sensors and communicates the relevant information to the PathoWARE platform. Moreover, the PathoSENSE IoT Gateway can also be connected to a SCADA service from the water utility authority, which will provide additional to the PathoSENSE sensors, real-time and historical telemetry data of both hydraulic and water quality dynamics, which are also essential for configuring the PathoINVEST simulations.

PathoTWEET is a cloud-based technology for monitoring anonymous data (text and photos) generated by citizens on social media platforms, that are relevant to the quality of water within the affected area. As a result, social media users can be considered "human sensors", which can assist in identifying the scale of a contamination event, and the source area, faster.

PathoVIEW is a set of technologies that enables First Responders to utilize smart devices on the field. For example, a First Responder may be receiving an alert on their smartwatch if they are entering a neighborhood that is receiving contaminated water, or they may use Augmented Reality glasses to overlay a map of the area that is being affected, as computed by PathoINVEST.

PathoGIS is a web application for real-time representation of geospatial data relevant to the emergency. It also serves as a visualization tool for the outputs of PathoINVEST, as it provides maps illustrating contamination evolution, as well as locations for sampling and sensor deployment.

PathoALERT is a system integrated with PathoWARE, that implements algorithms for real-time data analytics received from the PathoSENSE IoT Gateway, the PathoTWEET service, and the PathoVIEW smartphone app.

PathoTHREAT is a knowledge database with historical and scientific information on water contamination events and pathogen characteristics. The Incident Commander, after gathering information from similar emergency events, creates requests for the PathoINVEST. Moreover, PathoTHREAT will provide an API that will allow interaction with the knowledge database, and PathoINVEST will be able to use pathogen-specific models (e.g., using the EPANET-MSX data structure).

4 ILLUSTRATIVE CASE STUDY

The case study presented in this section focuses on demonstrating how PathoINVEST could be used during a contamination emergency, within the framework of PathoCERT, for forecasting the evolution of the event and identifying its source, to respond appropriately to the emergency.

The scenario utilizes the benchmark network "L-Town", created for the needs of the BattLeDIM (Battle of the Leakage Detection and Isolation Methods) competition [29] and it is based on a realistic network in Cyprus. The network has been suitably modified for security purposes. It is comprised of 782 junctions and 905 pipe segments, and it is assumed to provide water to around 10,000 citizens and industries. Each node has a randomly generated demand, synthesized from realistic data.

The L-Town network has two chlorination points at the inlets of the network, while the water supplied is assumed to have a constant concentration of Total Organic Carbon (TOC) of 1 mg/L. Water-quality dynamics, specifically the disinfection (chlorine) reactions in both the bulk and wall phase, as well as inactivation kinetics, are also incorporated in the benchmark. In the bulk phase, chlorine reacts with a series of reactants such as natural organic matter, and pathogens, whereas in the wall phase, chlorine reacts with biofilm. Finally, water quality parameters such as pH and temperature are also incorporated in the reaction model as they are factors that influence the chlorine demand and disinfection efficacy in the network.

We further assume that the PathoWARE service is already active in L-Town and that Standard Operating Procedures are already in place relevant to water contamination event management.

4.1 Emergency and establishment of Command & Control Center

An earthquake of 6.3 magnitude occurs near the city of L-Town. Damages on various buildings and infrastructures are being reported. The local authorities request the assistance of First Responders to set up a Command & Control (C&C) Center to manage the situation, coordinated by an Incident Commander. Following the Standard Operating Procedures, at the C&C Center, a dedicated area is assigned for collecting and managing information concerning water contaminations. The PathoGIS, PathoTHREAT, and PathoALERT are set up within the C&C Center, and PathoWARE connectivity is established.

4.2 Evaluating the risk

The Incident Commander requests an evaluation of the risk of possible waterborne contamination events due to the earthquake, using the PathoTHREAT tool. Similar past contamination events indicate that this is a likely scenario and that this can have a significant impact on the population within the next 12 hours.

The Incident Commander requests an evaluation of the situation from the utility operators, and both the drinking water and sewerage operators report abnormal pressure irregularities, which could be due to some leakage events. From past experiences, water and sewerage utilities have identified 5 “vulnerable” locations in their networks S = {S1, ..., S5}. At those points, there is no horizontal separation between water mains and sewer mains, resulting in an intersection of the two infrastructures with the former being below the latter. In the event of a severe earthquake, these 5 areas are potential sources of contamination due to wastewater leakage and subsequent infiltration into the DWN (Figure 2).

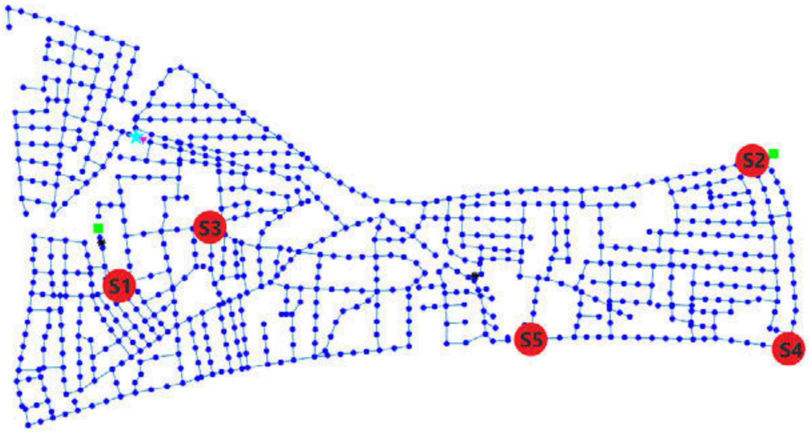

Figure 2. The 5 “vulnerable” areas within the network (S1-S5). The two reservoirs of the network are depicted as green squares and the tank as a cyan star.

As a result, the Incident Commander requests from the PathoINVEST operator, an impact assessment for all the different contamination scenarios (the 5 vulnerable locations) for the next 2-12 hours. Additionally, the Incident Commander requests maps that will show: a) contaminant evolution, b) optimal sensor placement, and c) sampling locations for source identification. For the contamination scenarios, single faults are considered as well as simultaneous (maximum two) faults with $\binom{5}{2} = 10$ maximum possible combinations. Although there are no results from sampling yet, it is suspected that wastewater has infiltrated the network, and therefore pathogen indicator *Escherichia coli* will be modeled in PathoINVEST. It is assumed that the contamination is continuous throughout the whole simulation and after exploring the literature, a conservative concentration of 3190 CFU/L *E.coli* and 140 mg/L (after 1% wastewater dilution) additional Total Organic Carbon (TOC) are modeled [30, 31]. Fast and slow reactions of chlorine with TOC and Chick-Watson inactivation kinetics are incorporated, with a pH of 7.5 and a steady temperature of 25°. Finally, no natural decay of *Escherichia coli* is included in the model, since this is insignificant in the time scale of the emergency event.

4.1.1 Forecast the evolution of possible contamination events

Figure 3 demonstrates the estimated *Escherichia coli* propagation through the network at 2, 4, 8, and 12 hours after the initial contamination using $S1$ and $S2$ as sources, whereas Figure 4 depicts the evolution of scenarios where there is a combination of simultaneous contamination sources $S3 + S4$ and $S4 + S5$ respectively. The PathoINVEST expert generates these maps (in an MP4 and GeoJSON format) and sends them through PathoWARE to the PathoGIS for visualization. Through this, the Incident Commander has a (near) real-time estimate of the possible propagation of contaminated water for the first 12 hours of each contamination scenario considered.

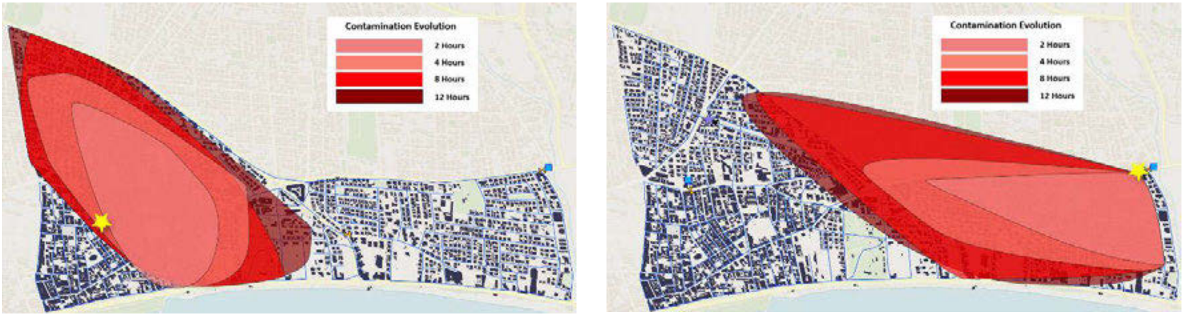

Figure 3. The contamination propagation for 2,4,8, and 12 hours when the source is S1 (left) and S2 (right).

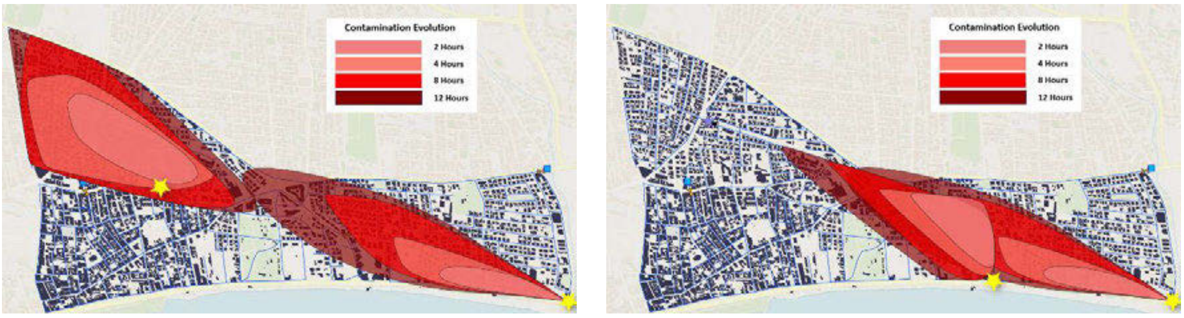

Figure 4. The contamination propagation for 2, 4, 8, and 12 hours for the combination of sources S3 + S4 (left) and S4 + S5 (right).

4.1.2 Impact assessment

The contamination Impact $I(k)$ in this case study is defined as the number of people affected by a contamination event, up until a specified discrete time k. Let the affected people be those who have ingested at least 1.0 Colony Forming Unit (CFU) of a microbiological contaminant mass until the considered time-step k. For this case study, we assume that the reference pathogen is *Escherichia coli* 0157. The impact is then calculated by simulating the reference pathogen concentration at each node of the network and calculating the total mass consumption at each node. The contaminant ingestion per person is calculated by considering the population estimate at each node, which is correlated to the base water demand at the node. The average water consumption in L-Town is assumed to be 150 L/person/day, where only 1% accounts for ingestion of tap water.

In PathoINVEST the operator is able to configure the contamination parameters and dynamics, using the pathogen reaction model extracted from the PathoTHREAT knowledge base (e.g., in an EPANET-MSX format), as well as other epidemiological parameters related to the pathogen, including the relevant exposure routes (i.e., ingestion, inhalation, dermal exposure). PathoINVEST can be used to compute an impact analysis for all 15 scenarios discussed previously (5 single plus 10 combinations), and the results are then communicated in a JSON format to PathoWARE, (an example can be found in Table 2). S1 and S2 have the highest impact and S5 the lowest.

Table 2. Impact of selected scenarios, defined as the percentage of affected people in the network.

	S1	S2	S3	S4	S5	S1&S2	S2&S3	S3&S4	S4&S5
2 hours	16.3%	9.7%	2.4%	1.8%	3.7%	26.1%	12.1%	4.2%	5.6%
4 hours	23.1%	19.4%	2.5%	3.2%	7.1%	42.5%	21.9%	5.7%	10.3%
8 hours	36.7%	32.9%	15.8%	9.2%	11.8%	69.6%	48.7%	24.9%	17.3%
12 hours	41.7%	37.4%	19.9%	18.7%	13.5%	76.8%	55.8%	38.1%	20.3%

4.3 Establish an early warning system for contamination

The Incident Commander reviews the impact assessment results, and following the recommendations from the expert operators of PathoTHREAT, decides to establish an early warning system by requesting the installation of PathoSENSE sensors within the DWN. The Incident Commander requests from the PathoINVEST operator to prepare a sensor placement map so that water utility staff together with First Responders can install and integrate them to PathoWARE through the PathoSENSE IoT Gateway.

For this illustrative case study, we assume that the water authorities are equipped with 3 PathoSENSE sensors for monitoring pathogens, which can be installed at any of the 20 sampling nodes (locations in the network that have been designed to allow water sampling and installation of mobile sensors during normal network operation). As a note, it is not possible to install mobile sensors everywhere in the network, due to physical and technological constraints.

As soon as the request reaches PathoINVEST operator, the Sensor Placement module is activated. To solve this problem, the tool needs to identify among the 20 possible sampling nodes, which 3 are the most suitable for monitoring a possible contamination event, considering that one (or more) of the 5 vulnerable locations initially identified may experience wastewater infiltration. For this problem, there are $\binom{20}{3} = 1140$ possible solutions. The goal is to identify the optimal combination of solutions (sensor locations), which minimizes the impact across all the possible contamination scenarios (i.e., all the combinations of the 5 potential sources). As these events have a low probability, the experts suggest considering up to two potential contamination sources. Moreover, it is assumed that the impact stops as soon as the contaminant is detected by any of the installed sensors, in the sense that mitigation actions could be taken after that point.

The sensor placement results are illustrated in Figure 5, with the 20 potential sensor nodes (left), as well as the selected sensor locations after performing the sensor placement analysis (right) — these {"413", "268", "206"}. The results are communicated in a GeoJSON format and are presented to the Incident Commander through PathoGIS. Instructions are then given to the appropriate teams for installing and setting up the PathoSENSE sensors as well as establishing communication links with the PathoSENSE IoT Gateway. Moreover, PathoALERT is configured to trigger an alarm, when the readings from these sensors deviate from normal.

Figure 5. The 20 possible locations for sampling and sensor placement (left) and the 3 best suitable locations for sensor placement (right).

4.4 Identifying the source

A few hours after the early warning system has been set up, one of the PathoSENSE sensors installed in the network ("206") measures high pathogen concentrations which trigger a PathoALERT notification. The Incident Commander issues a request for PathoINVEST to analyze the data to isolate the source. The PathoINVEST operator configures the Source Identification tool, to indicate which of the sampling locations are suspected to be contaminated (as determined by sensor readings, manual samplings, or consumer complaints).

The analysis uses the installed sensor measurements to identify a set of possible source locations of the contamination event. As already mentioned, there are 15 possible contamination scenarios. By running simulations for each scenario, a binary indicator is computed for each installed sensor, depending on whether they have detected a contaminant (1) or not (0). The binary signatures for all possible scenarios are illustrated in Table 3.

Table 3. The binary signatures for all possible contamination scenarios were generated using the installed sensor measurements. In gray, are the scenarios that match the sensor observations.

Scenario	S1	S2	S3	*S4*	*S5*	S1+S2	S1+S3	S1+S4	S1+S5	S2+S3	S2+S4	S2+S5	*S3+S4*	*S3+S5*	*S4+S5*
Sensor 1 (node 413)	1	0	0	***0***	***0***	1	1	1	1	0	0	0	***0***	***0***	***0***
Sensor 2 (node 268)	0	1	0	***0***	***0***	1	0	0	0	1	1	1	***0***	***0***	***0***
Sensor 3 (node 206)	0	1	0	***1***	***1***	1	0	1	1	1	1	1	***1***	***1***	***1***

Let the observed binary signature in this case study, using the 3 installed sensors, be $[\ 0\ 0\ 1\ ]^T$. Given the observed signature and after examining Table 3, a reduced set of possible contamination scenarios with the same signature can be identified, denoted by $S_r = \{S4, S5, S3 + S4, S3 + S5, S4 + S5\}$.

At this point, the PathoINVEST operator cannot distinguish which of the scenarios found in S_r is the actual contamination's source. For identifying the actual source, PathoINVEST needs more information, which can be acquired by requesting a team to perform manual sampling in the field. The sampling can be performed using a portable rapid testing sensor, such as PathoTeSTICK. It is assumed that any of the 20 accessible locations in the network (see Figure 5) are also suitable for manual sampling. Through simulation, the binary signatures are computed considering each scenario in S_r, for each possible sampling location (Table 4). According to Table 4, including the results from manual water sampling at locations {458, 662} to the existing binary indicators of available sensors {413, 268, 206}, will provide a unique signature to each of the scenarios in S_r. The PathoINVEST source identification tool would then instruct the operator to take samples, first from 662 and then from 458, if the algorithm is set up to give priority to True Positives rather than True Negatives.

Table 4. The 20 sampling locations and their respective expected signatures for the reduced set of contamination scenarios. Note that sensors already exist at nodes {413, 268, 206}.

	Available sampling locations (Node)																			
S_r	***413***	***268***	***206***	**458**	**662**	**333**	**347**	**275**	**12**	**764**	**607**	**738**	**218**	**484**	**410**	**454**	**135**	**571**	**25**	**83**
S4	*0*	*0*	*1*	0	1	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0
S5	*0*	*0*	*1*	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0
S3+S4	*0*	*0*	*1*	1	1	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0
S3+S5	*0*	*0*	*1*	1	0	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0
S4+S5	*0*	*0*	*1*	0	1	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0

The first field analysis is performed at node 662 using PathoTeSTICK, indicating that there are no traces of pathogen. This result is communicated back to PathoWARE and PathoINVEST is updated. Therefore, the contamination scenarios $S4$, $S3 + S4$, and $S4 + S5$ are removed from the set of possible contamination scenarios S_r. At around the same time, citizen complaints increase on social media and in the Water Utility customer service, complaining about the taste, color, and smell, in the northwest part of L-Town. Social media are monitored via PathoTWEET, thus PathoALERT issues the alarm for the increasing consumer complaints. This information is communicated to PathoINVEST from the Command & Control Center, and PathoINVEST suggests requesting another sampling at node 458 to validate the "human sensors. The sampling analysis confirms contamination at that node, and therefore, PathoINVEST reasons that the contamination scenario that is most likely to have occurred is the infiltration of wastewater in both $S3$ and $S5$ locations (Figure 6).

Figure 6. The final contamination map with all the necessary information.

Based on the verified source(s), the final contamination maps are computed and sent in GeoJSON format to PathWARE, to be then depicted through PathoGIS. Moreover, first responders on the field are receiving information whenever they are entering a contaminated area, or when the contamination has been extended to the area where they operate.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we propose an architecture for a decision support system (PathoINVEST), that aims to assist First Responders and Water Authorities in investigating and managing pathogen contamination events that could occur in a DWN after an emergency (such as an earthquake that causes damage to drinking water and sewerage infrastructure). An illustrative case study is described as a complete proof-of-concept, to demonstrate how the system could operate during a real emergency, considering the limitations in time and information. For this case study, simplified algorithms were implemented to demonstrate how each module can function. Our future work will investigate and propose new methodologies for modeling wastewater infiltration, as well as determining the most suitable mitigation and policies to minimize the impact of contamination. The case study highlighted that through the interoperability of the PathoCERT modules, a contamination event can be assessed and managed in a timely and effective manner.

PathoINVEST assumes that standard operating procedures are in place, to guide First Responders in establishing early warning systems and managing such an emergency event. Moreover, it assumes that the PathoWARE service is operational and that PathoSENSE sensors are available. Currently, all technologies described in this paper are under development and are expected to be evaluated and validated in field and tabletop pilot exercises by First Responders within 2023.

6 DATA AVAILABILITY

The data models and code generated and used for this case study are available in the following repository: https://github.com/KIOS-Research/PathoINVEST-WDSA-CCWI-2022

7 ACKNOWLEDGMENTS

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883484 (PathoCERT) and is supported by the European Union Horizon 2020 program Teaming under Grant Agreement No. 739551 (KIOS CoE) and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

8 REFERENCES

[1] T. Braeye, K. De Schrijver, E. Wollants, M. Van Ranst, and J. Verhaegen, "A large community outbreak of gastroenteritis associated with consumption of drinking water contaminated by river water, Belgium, 2010," Epidemiology & Infection, vol. 143, no. 4, pp. 711–719, 2015.

[2] J. Laine, E. Huovinen, M. Virtanen, M. Snellman, J. Lumio, P. Ruutu, E. Kujansuu, R. Vuento, T. Pitkanen, I. Miettinen, and J. Herrala, "An extensive gastroenteritis outbreak after drinking-water contamination by sewage effluent, Finland," Epidemiology & Infection, vol. 139, no. 7, pp. 1105–1113, 2011.

[3] F. Y. Ramirez-Castillo, A. Loera-Muro, M. Jacques, P. Garneau, F. J. Avelar-Gonzalez, J. Harel, and A. L. Guerrero-Barrera, "Waterborne pathogens: detection methods and challenges," Pathogens, vol. 4, no. 2, pp. 307–334, 2015.

[4] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, and C. Faloutsos, "Efficient Sensor Placement Optimization for Securing Large Water Distribution Networks," Journal of Water Resources Planning and Management, vol. 134, no. 6, pp. 516–526, 11 2008.

[5] R. Murray, T. Haxton, R. Janke, W. E. Hart, J. Berry, and C. Phillips, Sensor Network Design for Drinking Water Contamination Warning Systems: A Compendium of Research Results and Case Studies Using the TEVA-SPOT Software. EPA, 2010.

[6] A. Ostfeld, J. G. Uber, E. Salomons, J. W. Berry, W. E. Hart, C. A. Phillips, J.-P. Watson, G. Dorini, P. Jonkergouw, Z. Kapelan, F. di Pierro, S.-T. Khu, D. Savic, D. Eliades, M. Polycar-pou, S. R. Ghimire, B. D. Barkdoll, R. Gueli, J. J. Huang, E. A. McBean, W. James, A. Krause, J. Leskovec, S. Isovitsch, J. Xu, C. Guestrin, J. VanBriesen, M. Small, P. Fischbeck, A. Preis, M. Propato, O. Piller, G. B. Trachtman, Z. Y. Wu, and T. Walski, "The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms," Journal of Water Resources Planning and Management, vol. 134, no. 6, pp. 556–568, 2008.

[7] D. G. Eliades and M. M. Polycarpou, "A Fault Diagnosis and Security Framework for Water Systems," IEEE Transactions on Control Systems Technology, vol. 18, no. 6, pp. 1254–1265, 11 2010.

[8] L. Perelman, J. Arad, M. Housh, and A. Ostfeld, "Event detection in water distribution systems from multivariate water quality time series," Environmental Science & Technology, vol. 46, no. 15, pp. 8212–8219, 2012.

[9] D. G. Eliades, D. Stavrou, S. G. Vrachimis, C. G. Panayiotou, and M. M. Polycarpou, "Contamination Event Detection Using Multi-level Thresholds," in Computing and Control for the Water Industry (CCWI), vol. 119, no. 1, 2015, pp. 1429–1438.

[10] P. Fraternali, A. Castelletti, R. Soncini-Sessa, C. Vaca Ruiz, and A. E. Rizzoli, "Putting humans in the loop: Social computing for Water Resources Management," Environmental Modelling and Software, vol. 37, pp. 68–77, 11 2012.

[11] M. C. Getchell and T. L. Sellnow, "A network analysis of official Twitter accounts during the West Virginia water crisis," Computers in Human Behavior, vol. 54, pp. 597–606, 1 2016.

[12] S. Andreadis, G. Antzoulatos, T. Mavropoulos, P. Giannakeris, G. Tzionis, N. Pantelidis, K. Ioannidis, A. Karakostas, I. Gialampoukidis, S. Vrochidis, and I. Kompatsiaris, "A social media analytics platform

visualising the spread of covid-19 in Italy via exploitation of automatically geotagged tweets," Online Social Networks and Media, vol. 23, p. 100134, 2021.

[13] W.H.O., "Quantitative microbial risk assessment: application for water safety management," 2016.

[14] W.H.O, "Guidelines for drinking-water quality: first addendum to the fourth edition," 2017.

[15] P. Teunis, M. Xu, K. Fleming, J. Yang, C. Moe, and M. LeChevallier, "Enteric virus infection risk from intrusion of sewage into a drinking water distribution network," Environmental Science & Technology, vol. 44, no. 22, pp. 8561–8566, 2010.

[16] M. Blokker, J. Vreeburg, and V. Speight, "Residual chlorine in the extremities of the drinking water distribution system: the influence of stochastic water demands," International Conference on Computing and Control for the Water Industry (CCWI2013), vol. 70, pp. 172–180, 2014.

[17] M. Blokker, P. Smeets, and G. Medema, "Quantitative microbial risk assessment of repairs of the drinking water distribution system," Microbial Risk Analysis, vol. 8, pp. 22–31, 2018.

[18] N. Pelekanos, D. Nikolopoulos, and C. Makropoulos, "Simulation and vulnerability assessment of water distribution networks under deliberate contamination attacks," Urban Water Journal, vol. 18, no. 4, pp. 209–222, 2021.

[19] J. Schijven, J. M. Foret, J. Chardon, P. Teunis, M. Bouwknegt, and B. Tangena, "Evaluation of exposure scenarios on intentional microbiological contamination in a drinking water distribution network," Water Research, vol. 96, pp. 148–154, 2016.

[20] D. Jerez, H. Jensen, M. Beer, and M. Broggi, "Contaminant source identification in water distribution networks: A bayesian framework," Mechanical Systems and Signal Processing, vol. 159, p. 107834, 2021.

[21] F. Shang, J. G. Uber, and M. M. Polycarpou, "Particle backtracking algorithm for water distribution system analysis," Journal of Environmental Engineering, vol. 128, no. 5, pp. 441–450, 2002.

[22] D. G. Eliades and M. M. Polycarpou, "Water contamination impact evaluation and source-area isolation using decision trees," Journal of Water Resources Planning and Management, vol. 138, no. 5, pp. 562–570, 2012.

[23] S. G. Vrachimis, R. Lifshitz, D. G. Eliades, M. M. Polycarpou, and A. Ostfeld, "Active contamination detection in water-distribution systems," Journal of Water Resources Planning and Management, vol. 146, no. 4, p. 04020014, 2020.

[24] A. Seth, G. A. Hackebeil, T. Haxton, R. Murray, C. D. Laird, and K. A. Klise, "Evaluation of chlorine booster station placement for water security," in Computer Aided Chemical Engineering. Elsevier, 2019, vol. 47, pp. 463–468.

[25] F. Shang, J. Uber, and L. Rossman, "Epanet multi-species extension software and user's manual," U.S. Environmental Protection Agency, Washington, DC, EPA/600/C-10/002, 2011.

[26] D. E. Helbling and J. M. VanBriesen, "Modeling residual chlorine response to a microbial contamination event in drinking water distribution systems," Journal of Environmental Engineering, vol. 135, no. 10, pp. 918–927, 2009.

[27] E. W. Betanzo, R. Hofmann, Z. Hu, H. Baribeau, and Z. Alam, "Modeling the impact of microbial intrusion on secondary disinfection in a drinking water distribution system," Journal of Environmental Engineering, vol. 134, no. 4, pp. 231–237, 2008.

[28] A. Canciu, M. Tertis, O. Hosu, A. Cernat, C. Cristea, and F. Graur, "Modern analytical techniques for detection of bacteria in surface and wastewaters," Sustainability, vol. 13, no. 13, 2021.

[29] Vrachimis, Eliades, Taormina, Ostfeld, Kapelan, Liu, Kyriakou, Pavlou, Qiu, and Polycarpou, "BattLeDIM: Battle of the Leakage Detection and Isolation Methods," 2020.

[30] J. Yang, O. D. Schneider, P. K. Jjemba, and M. W. Lechevallier, "Microbial risk modeling for main breaks," Journal American Water Works Association, vol. 107, no. 2, pp. E97–E108, 2015.

[31] L. Metcalf, H. P. Eddy, and G. Tchobanoglous, Wastewater engineering: treatment, disposal, and reuse. McGraw-Hill New York, 1991, vol. 4.