IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 25, 2021, accepted December 30, 2021, date of publication January 5, 2022, date of current version January 13, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3140391

An Internet-of-Things Enabled Smart System for Wastewater Monitoring

FERNANDO SOLANO[1,2], STEFFEN KRAUSE[3], AND CHRISTOPH WÖLLGENS[3]

[1]Warsaw University of Technology, 00-665 Warsaw, Poland
[2]Blue Technologies sp. z o.o., 02-684 Warsaw, Poland
[3]Bundeswehr University Munich, 85579 Munich, Germany

Corresponding author: Fernando Solano (fs@tele.pw.edu.pl)

This work was supported in part by the European Union's Horizon 2020 Research and Innovation Program through the H2020 Micromole and H2020 SYSTEM Projects under Grant 653626 and Grant 787128.

ABSTRACT We present and evaluate an IoT-enabled sensing and actuating system for localizing illegal industrial harsh discharges of polluting wastewater in sewer networks. The special conditions of the sewer environment bring special challenges for the design of an IoT system and of its real-time algorithm for anomaly detection and localization in wastewater networks. The proposed design fulfills these requirements by using a new IoT architecture pattern, which we generalize and name Hop-by-hop Anomaly Detection and Actuation (HADA). The distributed anomaly detection and localization algorithm makes predictions over previous sensor measurements, while taking into account seasonality effects of wastewater and noise of the sensors. Based on simulations in a large network with three common illegal industrial wastewater pollutants, the advantages and limitations of the proposed wastewater anomaly localization system are discussed. The IoT system, including its anomaly detection and localization algorithm, was implemented using in a low-power microcontroller and tested in flowing wastewater with different harsh industrial waste.

INDEX TERMS Anomaly detection, sewage network, wastewater, waste water, IoT, design pattern, architecture pattern, Internet of Things, low-computer computing, micromole, Hop-by-hop anomaly detection and actuation.

I. INTRODUCTION

Discharges of sulfuric acid (H_2SO_4) to sewers could originate from applications, such as etching of semiconductors, accumulator acid or the production of organic chemical substances [3]. Sodium hydroxide (NaOH) is widely used for cleaning of surfaces in metal processing in industrial applications [4], whereas discharges of sodium sulfate (Na_2SO_4) can be caused by the regeneration of cation exchange resins, which are used for softening of water in industrial water treatment [5]. Illegal discharges of such dangerous harsh industrial waste into sewage networks could be harmful for the biological stage of WasteWater Treatment Plants (WWTP), its personnel, sewer pipes and the general public.

In order to detect such a dangerous discharge and to trigger remediatory actions promptly, continuous monitoring of wastewater characteristics is required.

Since not all illegal discharges of harsh industrial waste can be detected at the influent of WWTPs due to dilution effects, monitoring as close as possible to the point of discharge is necessary.

In this article, we present an IoT-enabled system - namely the Micromole system - for autonomous monitoring and sampling of harsh wastewater discharges in sewers' mainlines no smaller than 250 mm in diameter. The IoT system was initially designed and developed by a consortium of 11 organizations in Europe in the EU-funded project H2020 Micromole [1] and later improved in the EU-funded project H2020 SYSTEM [2]. The prototype is currently at Technology Readiness Level (TRL) 7 according to the European standards concerning research and innovation project outcomes.

This article presents the following contributions:

1) A new hardware architecture for wastewater monitoring IoT systems,
2) A distributed real-time algorithm for anomaly detection and localization of harmful wastewater discharges for constraint IoT devices,
3) A new IoT architecture pattern that can be reused in other applications and environments with similar constraints,

The associate editor coordinating the review of this manuscript and approving it for publication was Sayed Chhattan Shah.

TABLE 1. Notation table.

Symbol	Description
g	gravity constant
v	Velocity
d	Inner diameter of the sewer
J_E	Slope of the energy-line
k	Hydraulic roughness
ν	Kinematic viscocity
D	Dispersion coefficient
x	Distance from discharge point
σ_i	Standard deviation of the peak caused by dispersion at time i
h_i	Height of the peak caused by dispersion at time i

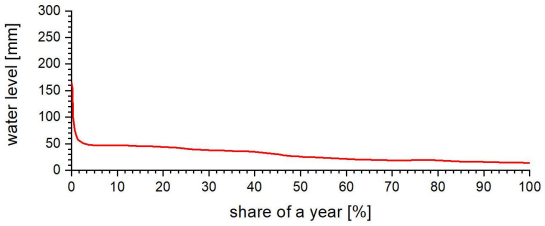


FIGURE 1. Cumulative Distribution of water level in a DN 350 sewer in the course of one year (typical hydrological conditions for central Germany).

4) Results from verification and validation of the proposed IoT solution through simulations and experimentation by potential end-users in a real wastewater environment.

This article is organized as follows. An overview of the phenomena involving wastewater advection, dispersion and diffusion in the sewer network is briefly explained in Section II. A summary of existing commercial devices and online algorithm for water anomaly detection is given in Section III. We provide the hardware design of the proposed IoT system devices and their electronic components - including sensor, actuators and radio communications - in Section IV. The distributed data processing algorithm used for automatic localization of harsh wastewater sources is described in Section V. We analyze the IoT system performance and its limitations, based on simulations with large network in Section VI. Finally, the prototype demonstration and validation activities carried out during the H2020 Micromole project with real flowing wastewater and harsh dangerous industrial waste are explained in Section VII.

II. BACKGROUND - WASTEWATER PROPAGATION AND OFFLINE SOURCE DETERMINATION

There are two types of sewer systems. *Combined sewers* are designed to drain both sewage and stormwater, whereas *separate sewers* are used to drain sewage only. As long as there is no stormwater flowing in combined sewers, the flow and the water level inside the sewers is usually very low as can be seen from Fig. 1. In such cases, the flow of wastewater is induced by gravity and the slope of the sewers. It can be described by the Colebrook-White-equation (1) for open channel flow:

$$v = -2\sqrt{2gdJ_E} \cdot \lg\left(\frac{k}{3.71d} + \frac{2.51\nu}{d\sqrt{2gdJ_E}}\right) \tag{1}$$

where v is the velocity, k the hydraulic roughness, J_E the slope of the energy-line, ν the kinematic viscocity and d the inner diameter of the sewer [6]. The notation used in this section is summarised in Table 1.

Dissolved substances, that shall be detected or sampled by the proposed IoT system are transported together with the flowing wastewater, which is called *advection*. In gravity sewers, a longitudinal stretching of solutes from areas with a high concentration to areas with a low concentration occurs which is refered to as *dispersion*. Dispersion is mainly caused by differences in the flow velocity across the cross section of the sewer. Furthermore, there is also a stretching of solutes by *Diffusion* caused by concentration gradients, which is, however, less important here. The entire process can be described by the advection-dispersion equation (2), where c is the concentration of dissolved substances and x the distance between two locations [7]. While the first term of this equation describes *advection*, the second describes the sum of *dispersion* and *diffusion*.

$$\frac{\partial c}{\partial t} + v\frac{\partial c}{\partial x} = \frac{\partial}{\partial x}\left[D \cdot \frac{\partial c}{\partial x}\right] \tag{2}$$

The discharge of a solute to the sewers will create a peak of concentration of this substance. The aforementioned process of dispersion is resulting in the broadening of such a peak and a decrease of its height. Turbulences, occuring in the manholes and other structures of a sewer system, will further increase dispersion [6], [8]. The concentration will be decreased also by the merge of multiple wastewater streams and the caused dilution.

The broadening of a peak in terms of its standard deviation σ, which is the peak width between its both inflection points, can be calculated as a function of the wastewater flow v, the distance (time t) between two points and the coefficient of dispersion D by the following equation:

$$\sigma_t^2 = \sigma_0^2 + 2D\frac{t}{v^2} \tag{3}$$

Here, σ_0 is the initial standard deviation of a peak and σ_t is the standard deviation of this peak after travelling the time t with the velocity v.

The coefficient of dispersion D represents the proportionality between a given gradient in concentration of a solute and in the velocity of flow and the resulting mass transfer between the respective volume elements. The value of D is influenced by properties such as the roughness of the sewer wall and the presence of sediments in sewers and manholes [9].

The mass of a solute can be expressed as the product of waste water flow and the concentration of the solute. The measurements of concentration over time or distance give a peak shaped signal. Based on the law of mass conservation it can by derived that the change of height of a peak is inversely proportional to the change of the peak width.

The area, A, under a symmetrical peak can be approximated by a triangle defined by peak height h and peak width at half peak height w_h, as follows.

$$A = w_h \cdot h \tag{4}$$

The error, caused by this approximation is in the range of 3% [10]. For non symmetrical peaks, there are similar approximations available taking into account the degree of asymmetry [11].

According to [10], the peak width at half peak height can be calculated using the following expression.

$$w_t = 2.354 \cdot \sigma \tag{5}$$

If we apply the law of mass conservation in terms of peak area at different points in time, we yield the following relationships:

$$A_t = A_0 \tag{6a}$$
$$w_{h,t} \cdot h_t = w_{h,0} \cdot h_0 \tag{6b}$$
$$2.354 \cdot \sigma_t \cdot h_t = 2.354 \cdot \sigma_0 \cdot h_0 \tag{6c}$$
$$h_t = \frac{\sigma_0}{\sigma_t} \cdot h_0 \tag{6d}$$

Combining equations (6d) and (2), we define the following relationship between the peaks' height:

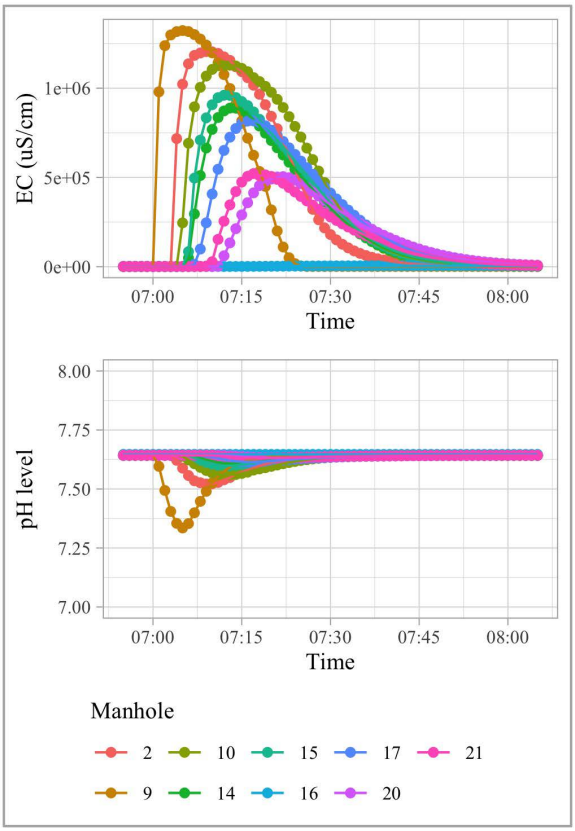
$$h_t = \frac{\sigma_0 \cdot h_0}{\sqrt{\sigma_0^2 + 2D\frac{t}{v^2}}} \tag{7}$$

The rate of change of the peak height with respect to the elapsed time can be inferred from (7) by taking the first derivative, as follows.

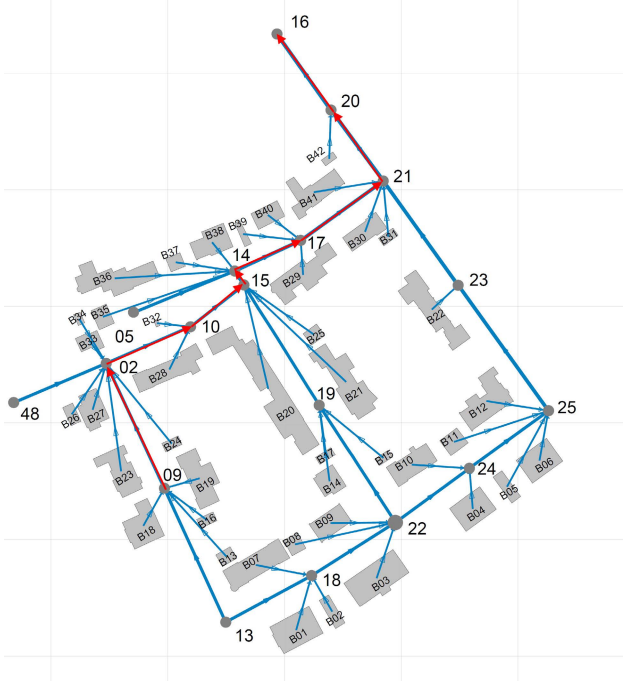
$$\frac{\partial h_t}{\partial t} = -\frac{\sigma_0 h_0 D}{v^2}\left[\sigma_0^2 + 2D\frac{t}{v^2}\right]^{-3/2} \tag{8}$$

Equation 8 shows that the rate of decay of the peak height over time is significant for discharges where the initial peak height magnitude is high. Nevertheless, discharges causing broad peaks at the source point or fast velocities in the sewers quickly reduce the decay rate of the peak height over time. As a result it is possible to observe fast changes in the concentration of the dissolved substance near its discharge point. We will exploit this fact in the design of our anomaly detection algorithm in Section V.

To illustrate this, based on numerical simulations of these hydraulic effects, Fig. 2a shows the changes in pH and Electrical Conductivity (EC) after an illegal discharge along the sewer network shown in Fig. 2b. Hydraulic characteristics of this network are summarized in Table 8. Fig. 2a demonstrates how the height - and shape - of pH and EC measurements would change for 9 sensors deployed along the flow path with a length of 274 m in total. As it can be observed, pH and EC signals caused by a polluting discharge tend to have a steep start and a long returning tail over time.



(a) EC and pH peak broadening and flattening caused by dispersion as seen by different manholes, when one hundred liters of sulfuric acid are discharged at location B13 of Fig. 2b in low wastewater flow conditions.



(b) Sewer network - Somewhere, EU. Red arrows indicate the flow path starting from manhole No. 09

FIGURE 2. Example of dispersion along a sewer network.

III. RELATED WORK

In this section we provide a survey of sensing systems for monitoring of sewers or related sensing systems that could be potentially adapted with little effort for monitoring wastewater.

A. SENSOR DEVICES

Sensor devices used to monitor different stages in mechanical, chemical and biological water treatment can also be used to measure sewage characteristics in manholes, if equipped with an appropriate mounting system. The WaterTechw^2 pH8000 Sensor [12] has been designed to provide pH and temperature measurements both for drinking water and wastewater applications. The sensor uses a flat surfaced electrode which includes an extended reference path, these features combine to provide a pH measurement, suitable for use in surface water, wastewater and drinking water applications. The electrode uses field proven flat surface and self-cleaning technology. The reference system is sustained by the Extended Path Reference design which provides a path to protect the reference in the presence of interacting ions such as proteins, silver and sulphides.

The AMC-1400 [13] is suited for wastewater applications as system for up to four sensor/transmitters, with a monitoring panel that provides vacuum fluorescent display for continuous gas concentration and alarm indication for each of four sensors. Housed in a fiberglass enforced polyester enclosure, the programmable AMC-1400 is designed to permit the user to configure relays, alarms and timers to their specific application. Engineering units are selectable from an internal memory library, including ppm, ppb, %L.E.L, and % volume. The AMC-1400 can be used with any of the array of hazardous gas sensor/transmitters of Armstrong Monitoring.

The Basic Ex 1 mobil sampler by ORI [14] includes the Mlog, a multi-parameter sensor and data logging system. It can also be equipped with sensors for pH/Redox, temperature and Conductivity. The Mlog is able to communicate with the sampler via Bluetooth or can trigger it by sending text messages when both devices are equipped with a GSM modem. The latter can also be used to send data to a cloud based monitoring centre.

In Table 2, we summarise most of the cyber-physical systems in the literature for monitoring or control of Water Distribution Systems (WDS) and Wastewater Monitoring, which we believe it could be of some application to the illegal discharge of polluting wastewater.

To the best of our knowledge, the IoT-enabled smart sensing system presented in this article - namely Micromole - is the first and only wireless sensing system capable of wastewater monitoring at any point within mainlines of sewer networks, aiding in the precise localization of a building discharging polluting wastewater illegally and collecting physical samples for further analysis.

B. ANOMALY DETECTION ALGORITHMS BASED ON WATER QUALITY PARAMETERS

Table 3 summarises the most important anomaly detection algorithms using water quality parameters that could be potentially adapted for wastewater monitoring. We make use of the nomenclature presented by Cook *et al.* in [23] concerning anomaly detection using IoT systems.

The sensing systems proposed by Rekha *et al.* and Saetta *et al.* in [24] and [21], respectively, use alarms triggered by a distance-based anomaly detection algorithm, with predefined fixed thresholds as the basic method for classification of a time-series of measurements as an anomaly. Such methods may perform properly for drinking water distribution networks and the monitoring of physical parameters of excreted urine, since no fluctuations in those physical parameters are expected in those applications at any time or point in the network in a normal context, *viz.*, no data seasonality. In contrast, our anomaly detection algorithm is designed as to adapt the baseline to the normal fluctuation of wastewater characteristics - such as EC and pH - during the day, therefore, considering the seasonality of wastewater fluctuations.

Zhang *et al.* in [27] propose the Dual Time-Moving Windows Anomaly Detection algorithm, which extends the Anomaly Detection and Anomaly Detection and Mitigation algorithms of Hill and Minsker in [28]. In a nutshell, these three algorithms keep a buffer of the past N measurements in order to make an estimation of the next measurement. A prediction is updated by considering the current prediction, the new measurement and a learning rate parameter, namely α, which is constant. As such, the running time for processing a new measurement is $O(N)$. As the learning rate is constant during the algorithm execution, the algorithm cannot adapt its learning rate based on the gained confidence from previous estimations. The anomaly detection algorithm presented in this paper has a running time complexity of $O(1)$.

The SIMONA sensor network nodes - refer to [17] and its reference in Table 2 - have implemented in-built clustering anomaly detection algorithms, proposed by Salvato *et al.* in [26]. The immune based algorithms comprises two consecutive phases in its operational lifetime: a) a learning phase - where the DBSCAN algorithm [29] is used for clustering the first 200 measurements - and, b) a classification phase - where new measurements are classified as normal or abnormal and the knowledge about the system is updated using immune based rules. DBSCAN is a clustering algorithm that partitions a set of N samples such that in each partition every sample has at least *minPts* samples in a radius of ϵ, where *minPts* and ϵ are parameters of the algorithm. Given N samples, the DBSCAN algorithm has a running complexity of $O(N)$ and a requirement in the order of $O(N)$ for memory space. Once clusters are formed by DBSCAN, the algorithm of Salvato *et al.* classifies a new measurement by including it to the closest cluster and updating immune based metrics of the chosen cluster. Before processing a new measurement,

TABLE 2. Different CPS propose in the literature for monitoring of water distribution systems and wastewater networks.

Ref.	Water Sensors[a]	Air Sensors[b]	Communications	Energy	Application	Mounting
[15]	T	-	-	-	Illegal sewage connections	mainline
[16]	T	-	-	-	Illegal sewage connections	mainline
[17]	COD, NH_3, TSS, WL, pH, T	H_2S	-	-	Wastewater Monitoring	mainline
[18]	-	-	-	GAC-SCMFC	WWTP Monitoring	network only
[19]	pH, T	VOC, T, RH	Ethernet	-	WDS Monitoring	open channels
[20]	P, F, pH, EC, T, ORP	-	GSM (3G)	33Ah	WDS Monitoring	pressure lines
[21]	pH, EC	-	Ethernet	-	Nonwater Urinals Control	plumbing
[22]	WL	-	GSM/GPRS	Battery of 32Ah	Level of water bodies	over water level

[a]T = temperature, TSS = Total Suspended Solids, WL = Water Level, COD = Chemical Oxygen Demand, Tu = Turbidity, TDS = Total Disolved Solids, F = flow, ORP = Oxidation Redox Potential

[b]VOC = Volatile Organic Compounds, T = temperature, RH = Relative Humidity

TABLE 3. Literature on anomaly detection algorithms based on water quality parameters.

Ref.	Algorithm type	Medium	Running time	Memory space	Baseline calculation	Anomaly classification
[21]	Threshold-based	Urine	$O(M)$	$O(M)$	Fixed	Threshold
[21]	Lasso regression model	Urine	$O(M)$	$O(M)$	Regression model	Threshold
[24]	Threshold-based	Wastewater	$O(1)$	$O(1)$	Fixed	Threshold
[25]	DBSCAN	Drinking water	$O(N)$	$O(N)$	DBSCAN	Noise of DBSCAN
[26]	Immune-based	Wastewater	$O(N)$	$O(N)$	DBSCAN	Immune-based
[27]	Dual-time Moving Windows	Waterbodies	$O(M)$	$O(M)$	Liner combination of M samples	t-distribution

the immune based algorithm may perform re-clustering of the samples and eliminate overlapping clusters. As a result, during the classification phase, the online processing of each sample has a worst-case running complexity of $O(N)$ and requires $O(N)$ of memory resources, where N is the number of samples. DBSCAN is also used in [25]. In our opinion, the immune base algorithm of Salvato *et al.* not only detect abnormal measurements, but may also have the capability of identifying the type of chemical substances in a measurement, which unfortunately was not evaluated by Salvato *et al.* in [26]. In our work, we provide an anomaly detection algorithm of lower running and memory complexity without the possibilities of substance identification.

Another large set of works not included in Table 3 use machine learning algorithms. Support-Vector Machine (SVM) approaches - as discussed in [30]–[34] - have difficulties in detecting a gradual anomalous change of sensor values in a time-series, as discussed in [30]. Other works - e.g., [33], [35]–[37] - use Artificial Neural Networks (ANN) for anomaly detection. In most cases, such machine learning algorithms are trained in the cloud, and then run for classification on the IoT edge nodes. The execution of machine learning classifiers in IoT resource-constrained devices should be still carefully evaluated for energy demands and time execution.

IV. HARDWARE ARCHITECTURE

The location of a harsh wastewater pollutant source can only be determined with all certainty by monitoring the outgoing wastewater from its private sewer pipe. Unfortunately, private sewer pipes are too small (with diameters below 120mm) for fitting a sensing device. Moreover, access to a private area – even in the sewage network – is allowed only after permission from the owner or a warrant issued by legal authorities. In either case, organizations discharging illegal waste would realise that their illegal activities will be monitored and take measures to avoid harmful discharges being detected.

The solution presented in this article aims at indirectly detecting such harmful discharges originated at a private inlet by monitoring the changes of the composition of wastewater in the sewer mainline before and after the private sewer pipe inlet, as illustrated in Fig. 3. The IoT-enabled monitoring system presented in this article - namely *Micromole* - is designed to work in such a way.

Our IoT system is composed of three IoT devices mounted in the mainline, and one IoT gateway device mounted at a manhole, below its lid - as observed in Fig. 3 - Cloud Services and an end-user application.

The construction of the hardware components for each one of our IoT devices is modular: sensors and major electronic components are isolated in their own housing, called *modules*, each one providing different functionalities. Each module has the same dimensions and a standardized wired interface for wired communication with other modules residing in the IoT device. The housings are streamlined as to reduce the probability of sewer pipes clogging due to the flow of large solids in wastewater. The composition of each IoT device in terms of its modules is summarised in Table 4.

A. LOCALIZATION OF POLLUTING SOURCES IN SEWER MAINLINES

The main operation principle used by Micromole for localization of polluting sources in wastewater is the following. The first two IoT devices (on the left part of Fig. 3) contain a pH and EC sensor each. The first IoT device is positioned before the private inlet that needs to be monitored, while the second device is positioned just after the monitored private inlet. The sensors located in these two IoT devices monitor

TABLE 4. Micromole IoT devices and their physical modules.

No.	Device name	pH sensor	EC sensor	Sampling	Radio Communication	Optical Communication	Mobile Communication
1	Reference	YES	YES	NO	YES	NO	NO
2	Master	YES	YES	NO	YES	NO	NO
3	Sampler	NO	NO	YES	YES	YES	NO
4	Gateway	NO	NO	NO	NO	YES	YES

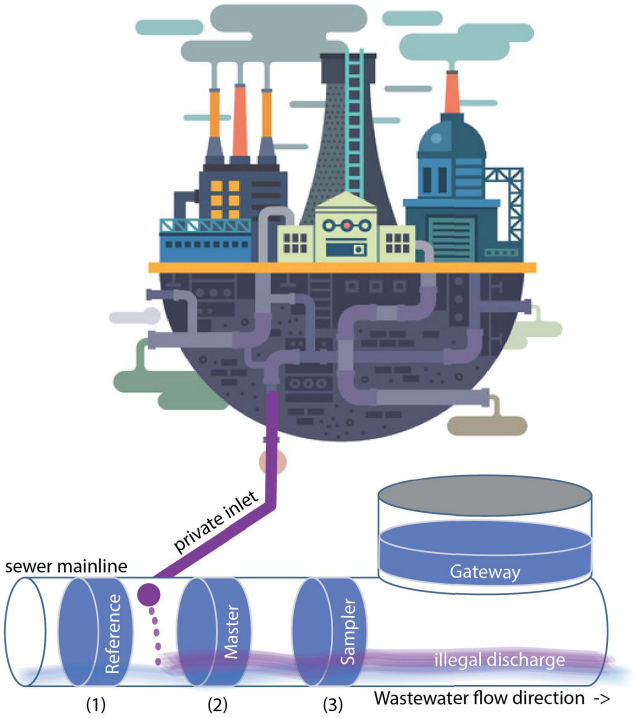


FIGURE 3. The Micromole system is a Wireless Sensor and Actuator Network (WSAN) composed of three IoT devices and an IoT gateway operating in sewer lines.

FIGURE 4. Distribution of sewer diameters for a medium-sized city.

the level of pH and EC every second on average. A significant difference in the values in pH and EC between the two IoT devices indicate the inflow of a suspicious wastewater coming from the monitored private inlet. More specifically, if the second IoT device notices an abrupt and significant change of pH or EC which was not observed by the first device, there is a strong indication of a discharge of harmful wastewater from the monitored private inlet. We refer to such approach as *anomaly localization via spatial correlation*.

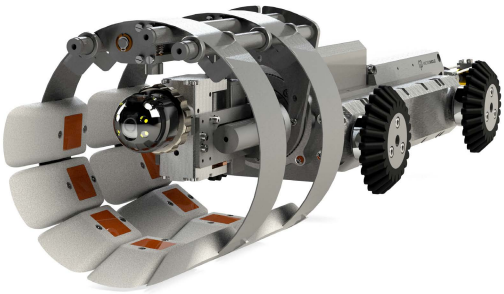
In case of an anomaly being detected and localized, the third IoT device is immediately activated. The third IoT device contains a sampling module. The sampling module is capable of pumping and storing samples of flowing wastewater in internal containers of approximately 2.5mL of volume each, which is sufficient for high sensitive analytical techniques such as Liquid Chromatography with tandem mass spectrometry (LC-MS-MS) or Gas-Chromatography with mass spectrometry (GC-MS). One sampling module contains three of such containers, each one with its own separate fluidic system for preventing sample cross-contamination. Wastewater samples can later be extracted from each container of the sampling module in the third device for laboratory for analysis. A description of the first version of the sampling module can be found in [38].

B. REMOTE INSTALLATION AND EXTRACTION OF MICROMOLE IoT DEVICES IN SEWER PIPES

A wastewater network is composed of sewer pipes of different diameters. The diameter of the target pipes for the Micromole system affect the dimensions of the IoT devices. We selected DN250 as the minimum acceptable pipe diameter for the IoT device after an analysis of the sewage pipe diameter distribution in a chosen city. Fig. 4 shows the distribution of the pipe diameters in that city, which hosts 300.000 inhabitants. The proposed sewage monitoring system is capable of fitting in such small sewage mainlines without blocking the wastewater flow.

Due to the small sewer pipe diameter, the Micromole IoT devices are designed as to be installed by a robot in the mainline. The company Pipeferret, Reykjavik, IS, has modified one of its crawler robots - normally used for sewage mainline inspection and industrial services such as pipe rehabilitation by inliners - for carrying and installing the sewage monitoring devices. Such installation can be performed remotely even from a manhole that is located 500 meters away from the desired point of monitoring. Fig. 5a shows the robot that was constructed together with two Micromole IoT devices.

Micromole IoT devices do not occupy more than 10% of the cross-section area of a 250DN pipe, in order to prevent sewage flow blocking and also to allow the transportation of the monitoring device by the robot. As a consequence, the thickness of the IoT monitoring devices shall not be larger than 3 centimetres.



(a) Micromole Crawler Robot for installation with two Micromole devices, where each IoT device is composed of five modules.

(b) Two micromole devices being installed by a robot in a DN250 pipe through a manhole in Legionowo, Poland, on February 2019.

FIGURE 5. The Micromole system and its crawler robot.

The thickness of the IoT monitoring device is of extreme importance for the following two factors:

1) it defines the inner volume allowed for electronic components, representing an important constraint in other system parameters, such as: battery capacity, type of sensor technology used and telecommunication technology used, which are presented below.
2) To monitor water quality in combined sewers, the height of the modules has to take into account the average low water level that is present at dry weather conditions. Fig. 1 presents water level statistics of a DN350 pipe.

C. SENSORS, TELECOMMUNICATIONS AND DATA PROCESSING MODULE

The first two IoT devices must continuously monitor the wastewater properties. For this, each one of these two devices is equipped with a pH and conductivity sensor, each one of these sensors is encapsulated in its own module. In [39], the reader can find a description of these two sensors and their housing for operating in the sewage environment.

Every IoT device contains a module for data gathering from sensor modules, data processing and information transmission, which we refer to as the *main* module in the Micromole IoT device.

Since the housing for the main was streamlined and its volume is small, commercially available boards - such as Arduino or Raspberry Pi - cannot be used. Instead, we opted for designing our own Printed Circuit Board containing only the necessary electronic components for energy efficiency. The main module hosts a low-power controller, a radio transceiver and a 2Gb flash memory. The main module executes the algorithm described in Section V.

The controller is a ARM Cortex-M4F controller running at 48MHz clock frequency and is able to query data from other modules using the IoT device bus. The controller is programmed with the RIOT operative system [40].

In order to avoid clogging of the sewer pipes, the Micromole IoT devices should avoid protruding cables. Therefore, an emphasis was made on providing wireless communications between the Micromole devices. The radio propagation conditions in an underground DN250 sewage pipes were evaluated with both a) a signal generator operating at different frequencies and a radio spectrum analyzer, and b) different battery operated portable transceivers operating at different unlicensed frequencies in Europe: 2.4 GHz (WiFi), 868 MHz (LoRa), 439 MHz (LoRa) and 169 MHz (FSK modulation). The best overall signal was achieved with the 169 MHz frequency due to its robustness against sewer humidity. On average, the maximum distance between the devices cannot exceed 10 meters when a +20dBm amplifier is used for the 169 MHz radio, requiring a maximum of 250mW of electrical power during transmission. Fig. 6 shows the received signal power for the 169 MHz radio.

A special radio driver was implemented to allow transmitting the 802.15.4 frames over the 169MHz radio. Thanks to this driver and the CoAP [41] and 6LoWPAN [42] libraries of RIOT-OS, the three Micromole IoT devices and its gateway establish an autonomous 6LoWPAN Wireless Sensor and Actuator Network (WSAN). In order to reduce energy consumption, sensor measurements are compressed using the algorithm described in [43].

In order to allow installing the system from a distant manhole, the third Micromole device is equipped with an optional optical communication module, allowing communications for 100 meters using a plastic optical fiber with the Micromole IoT gateway.

The Micromole IoT gateway is equipped with a 4G mobile communication module in order to relay data and alarms from the three devices in JSON format.

D. POWER AND BATTERY

Since there is no access to the power grid within the sewer network, all micromole IoT devices and its gateway must powered by batteries.

A 2600mAh battery is stored within a battery module. To extend the lifetime of the system several battery modules can be attached to one device, provided enough physical space at the device.

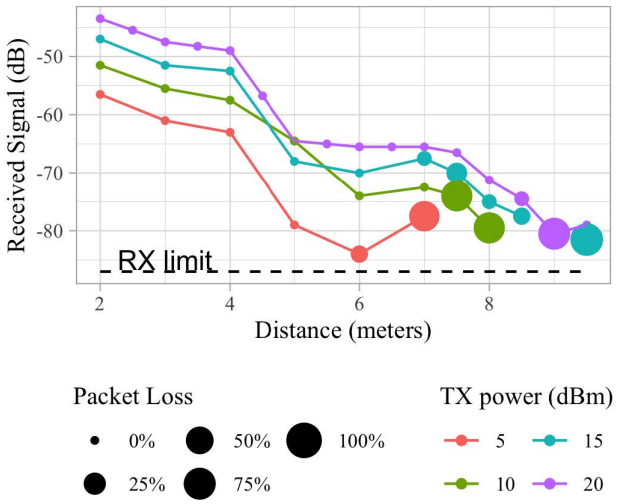


FIGURE 6. Received signal power in an DN250 sewage pipe three meters below ground level.

TABLE 5. Estimated energy consumption of Micromole IoT device major components per cycle of operation.

State	Component	Time (ms)	Current (mA)
Active	(new) EC	10	18
	pH	10	10
	Radio RX	4	10
	Radio TX	51.2	50
	Microcontroller	200	10.8
Idle	(new) EC	990	4
	pH	990	5
	Radio RX (sleep)	944.8	1.8
	Radio TX	0	0
	Microcontroller	800	3.1

A Micromole IoT device takes a EC and pH sample and processes them, as described in this article, every second. Every 10 seconds, a keep-alive frame is sent to the neighbouring devices. The currently implementation of the Micromole IoT devices allows powering of one device for 25 hours with one battery module. Most of the energy is currently consumed by the EC sensor measurements and radio transmissions.

The authors are developing a new EC sensor of lower energy consumption, which will be presented in a separate publication. In parallel, we are implementing the 802.15.4 non-beacon enabled sleep modes. Table 5 shows an estimation of the time and energy used per component when considering the new EC module and radio sleep modes.

V. DISTRIBUTED ANOMALY LOCALIZATION ALGORITHM OF HARMFUL WASTEWATER DISCHARGERS

In this section, we provide details about the distributed algorithm used for our anomaly localization approach.

A. DISTRIBUTED IoT DEVICE-BASED PROCESSING

Following the diagram of Fig. 3, we hereinafter name the first (left-most in the figure) monitoring IoT device - placed before the monitored inlet - as the *reference device*. The second monitoring IoT device - placed after the monitored inlet - is named the *master device*. The third monitoring IoT device (right-most in the figure) is named the *sampler device*.

As mentioned before, determining whether the harmful wastewater originates from a particular monitored household involves the comparison of two multivariate time-series of observations, one provided by the reference device and the second one provided by the master device. One IoT design solution for this application could consist on using a Fog or Edge computing approach [44] or Cloud approach, where all measurements from both sensors at each device are transferred to a server in the Fog or Edge for data processing and decision taking. Unfortunately, such an approach has the following disadvantages for our application:

1) high energy consumption due to the necessity of all samples transmissions in difficult radio propagation conditions in underground sewer pipes (refer to Section IV-C),
2) potential delays concerning the systems' time-critical decisions due to disruptions in the communication due to fluctuating wastewater level, and passing objects near the manhole's lid.

With respect to the second point above, time-critical decisions for the Micromole IoT system include triggering the sampling module at the correct time: during the time the illegal wastewater is flowing through the Micromole IoT devices. If not performed promptly, the sampler device would collect and fill containers without the targeted wastewater flow. Due to similar reasons, unfortunately, approaches for self-adaptive data sampling reduction - such as the one presented by Botero-Valencia *et al.* in [45] - may yield similar outcomes.

Instead of using a Cloud-center or Fog-center design pattern [46] for online anomaly detection, *we opted for designing a real-time algorithm by distributing the logic of analysing the signal between both IoT devices in a way that minimizes the number of data frames that are exchanged between both IoT devices*, hence, reducing overall energy consumption and bounding the response-time.

Hosting the anomaly localization algorithm on the IoT device implies the careful design of scalable resource-constrained algorithms that work in real-time. Fig. 7 presents the software architecture used for such functionality distribution. Each one of the IoT devices have implemented a *sensor driver* module that periodically queries the attached pH and EC sensor for a measurement through the device bus. Each measurement is collected by the *anomaly detection* algorithm within each IoT device for detection of abnormal suspicious patterns in the last sequence of measurements. If the anomaly detection algorithm at the reference device detects an anomaly, an alarm is sent to the *source determination* algorithm located at the master device. If the anomaly detection algorithm in the master device detects an anomaly, the alarm is passed internally to the source determination algorithm (which is located in the same device). The source

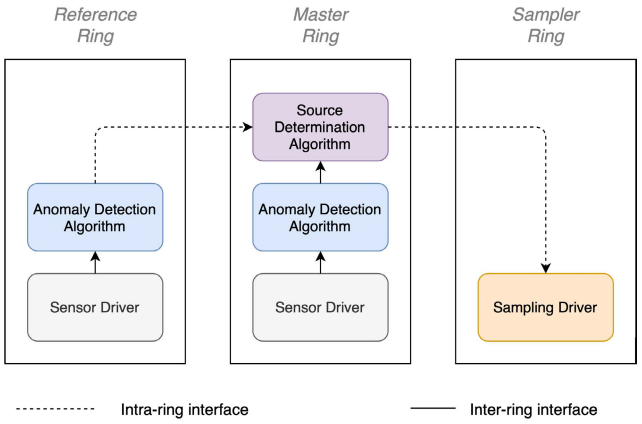


FIGURE 7. Software architecture of the distributed anomaly localization algorithm of Micromole for wastewater networks.

determination algorithm is in charge, therefore, of comparing and correlating both alarm notifications (if both are present) in time and magnitude in order to determine whether a harmful waste was disposed from the monitored inlet or before. In this way, *only the alarm indication needs to be passed between devices, without the need of transmitting all measurements*. The design of each one of these two algorithms is summarized in the remaining subsections.

B. ANOMALY DETECTION ALGORITHM

Anomaly detection is only possible if contextual information is properly defined. Defining a normal context is complex for wastewater sewage networks. It shall be considered that pH and EC measurements are *seasonal* [23]: the level of pH and conductivity of aggregated wastewater in the sewage network changes slowly during the day in normal conditions. As it can be observed in Fig. 8, the pH level and EC fluctuate continuously during the day, caused by discharging waste by users of sewage system. During the night, there is usually no wastewater flow and both physical characteristics drop.

Due to this intrinsic fluctuation of the pH and conductivity levels in the sewage, the anomaly detection algorithm proposed in this article is *predictive* [23]. The anomaly detection algorithm calculates and constantly updates a *baseline* used for prediction, which are calculated using a Kalman filter. Thresholds for triggering alarms are set relative to this baseline.

An outline of the anomaly detection algorithm used in our solution can be seen in Algorithm 1.

We assume that there is no correlation between the baseline values of pH and EC in time. As a consequence, our solution uses two independent Kalman filters: one for calculation of the baseline for pH, and another one for the EC baseline calculation.

We employ two-dimensional Kalman filters with the aim of inferring not only the value of the measured EC (or pH), but also the linear increase or decrease over time of these measurements. We assume that ageing of the sensor, clogging

Algorithm 1 Real-Time Anomaly Detection Algorithm

Input: A measurement value y_t
Output: alarm state z

if $x_t \geq H_1$ or $x_t \leq L_1$ **then**
  Restart filter state
  return no water
end if
Predict value of x_t (baseline)
Set $e \leftarrow x_t - y_t$
if $|e| < K$ **then**
  Correct Kalman filter using measurement x_t
else
  if $e < L_2$ or $e > H_2$ **then**
    Notify Source Determination Algorithm
    return anomaly detected alarm
  end if
end if
return no alarm

and bio-film (due to the permanent contact of the electrodes with wastewater) may cause a *drift* [23] in the observed values over time, for which we must compensate.

Our state vector is defined as $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, which indicates the expected level of pH (or conductivity) in the first component, x_1, and the linear rate of change of the first component in the second component, x_2. Besides the state vector $\boldsymbol{x}$, Kalman filter also keeps a covariance matrix, P, which provides an estimate of the accuracy of the estimation of the state vector.

Let $\boldsymbol{F}$ be the state transition matrix of our process, defined as follows:

$$F = \begin{bmatrix} 1 & \frac{1}{h} \\ 0 & 1 \end{bmatrix}, \quad \text{where} \tag{9}$$

h is the measurement rate of the sensor.

Let $\boldsymbol{Q}$ and $\boldsymbol{R}$ be the co-variance matrices of the process and observation noise, respectively. For our application, they are defined as follows:

$$Q = \begin{bmatrix} q_{11} & 0 \\ 0 & q_{22} \end{bmatrix} \tag{10}$$

$$R = r \tag{11}$$

where r is the variance of the EC (or pH) over time in normal conditions assuming a perfect sensing process, q_{11} is the variance of the noise of the EC (or pH) sensor signals, and q_{22} is the variance of the rate of change of EC (or pH) over time.

We estimate the value of r as the variance of the EC (or pH) values of the wastewater using high-precision laboratory equipment. The value of q_{11} and q_{22} are estimated by measuring the variance of the signals given by the Micromole EC (or pH) sensor in a solution of known and stable EC (or pH). The values of $\boldsymbol{F}$, $\boldsymbol{Q}$ and $\boldsymbol{R}$ are parameters of a Kalman filter and, hence, constant during its execution.

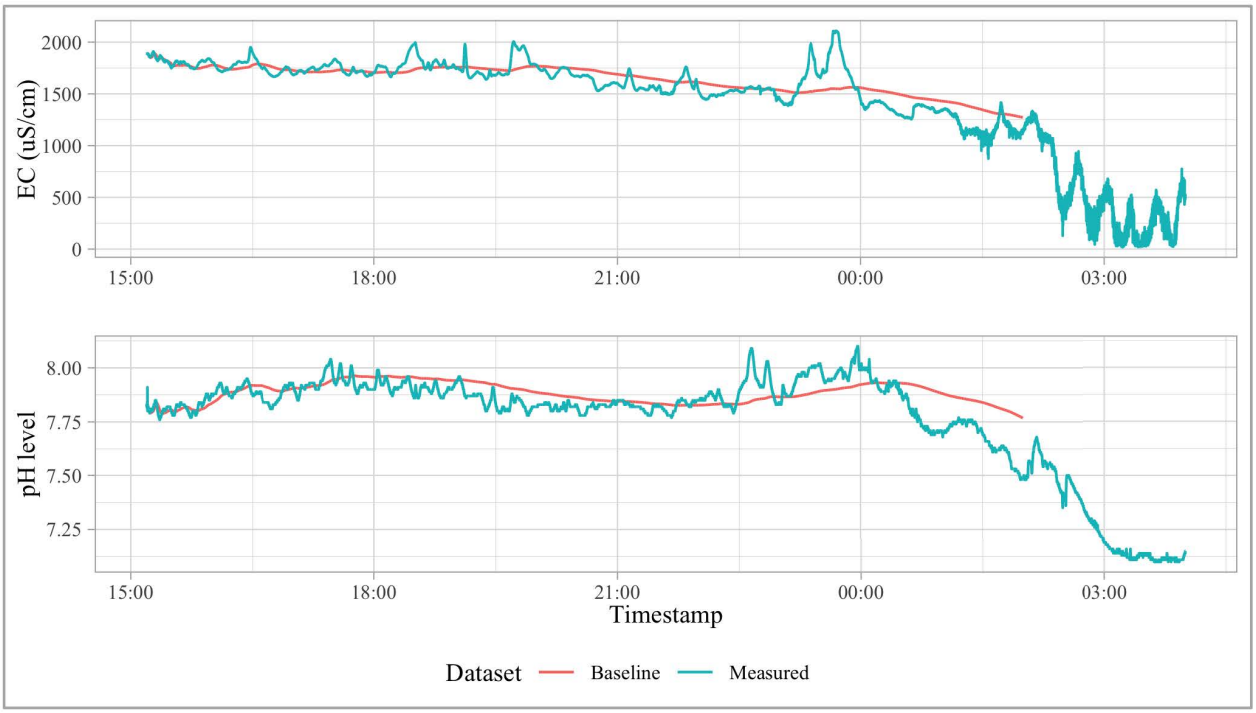


FIGURE 8. EC and pH measurements in a DN900 sewer pipe from Zissenbach, DE. Sept 9th to Sept 10th 2020. The sampling frequency for this experiment was 1*Hz*. Calculated baseline indicated as a red line.

Kalman filters have two phases: a prediction phase and a correction phase. In each phase the state variables and the covariance matrix, *viz.* $\boldsymbol{x}$ and $\boldsymbol{P}$, respectively, are updated in constant running time.

For our application, the prediction phase computes a new estimation of the state variable, named $\hat{\boldsymbol{x}}$, and of the covariance matrix as follows:

$$\hat{\boldsymbol{x}} = \boldsymbol{F}\boldsymbol{x} \tag{12a}$$

$$\boldsymbol{P} = \boldsymbol{F}\boldsymbol{P}\boldsymbol{F}^T + \boldsymbol{Q} \tag{12b}$$

Once a new measurement z of pH (or EC) is collected, the correction phase of the Kalman filter updates our state variables and covariance matrix as follows.

$$\boldsymbol{K} = \boldsymbol{P}(\boldsymbol{P} + \boldsymbol{R})^{-1} \tag{13a}$$

$$\boldsymbol{x} = \hat{\boldsymbol{x}} + \boldsymbol{K}(z - \hat{\boldsymbol{x}}) \tag{13b}$$

$$\boldsymbol{P} = (\boldsymbol{I} - \boldsymbol{K})\boldsymbol{P} \tag{13c}$$

The variable $\boldsymbol{K}$ is named the Kalman gain, which can be seen as a learning rate. The Kalman gain is calculated based on the covariance matrix and the observation noise. The Kalman gain weights how much of the previously predicted estimate, $\hat{\boldsymbol{x}}$, versus how much of the new provided measurement, z, should be considered during correction. The covariance matrix is updated at the end by taking into account the Kalman gain. In contrast to the learning rate of the algorithms proposed in [27], [28], the learning rate of our algorithm takes into account the confidence of the algorithm in previous predictions.

As a result, for estimating the pH (or EC) values, the Kalman filter requires a total of six variables to be kept between calls: two for the state vector and four for the covariance matrix. In contrast to the algorithms proposed in [26]–[28], neither the run time nor the memory usage of our algorithm scale with the number of previously considered samples, making it more suitable for IoT resource-constraint devices.

C. PAUSING AND RESUMING CORRECTIONS

In normal conditions, the baseline calculation is able to predict the upcoming measurement of EC (or pH), while compensating for process and observation *noise* [23]. However, if an illegal discharge of a substance with extreme values of pH or EC is discharged for a large period of time, the Kalman filter may improperly correct the Kalman state variables. In order to overcome such an effect, we include thresholds (constant K in Algorithm 1) for bypassing the current measurement to the Kalman filter correction phase. If the difference between the predicted value of the Kalman filter and the current measured value (variable e in Algorithm 1) is above this threshold, the correction phase is not done; allowing the baseline to continue the previously calculated trend (ignoring the measured value for baseline update). In such situations, we say that the Kalman filter is *paused*.

Once a measurement that is closely enough to the predicted value by the baseline is obtained, the correction phase of the Kalman filter is done. In such situations, we say that the Kalman filter was *resumed*.

In situations where the measured sensor signal is very close to the threshold and the noise of the sensor, the algorithm may toggle the Kalman filter states continuously. In order to prevent such undesired effect in the algorithm, we include a hysteresis control: the threshold for pausing the Kalman correction step - named *rising threshold* - is different from the threshold for resuming the Kalman correction step - named *falling threshold*. This is not included in the pseudo-code of Algorithm 1.

Not all changes in pH and EC may indicate the disposal of harmful waste. For instance, the regeneration of a dishwasher or the use of pipe cleaners may alter the level of EC and pH in wastewater, but are allowed to be discharged in the sewer systems. In such cases, it is desired to avoid corrections of the Kalman filter baseline while preventing the system from mistakenly identifying an anomaly. The sought illegal discharges are characterised by an extremely high change in pH and EC levels. Therefore, a second set of threshold values (namely L_2 and H_2 in Algorithm 1) must be assigned in order to be able to detect such harmful discharges, and discern them from legal occasional legal discharges. These thresholds are called *anomaly thresholds*. These thresholds are relative to the baseline. The thresholds for local alarms must be set higher than the threshold for pausing and resuming the Kalman filter correction phase.

As it can be seen in the last measurements from Fig. 8, the wastewater level may drop below the minimum acceptable level for the electrodes. If the water level in sewage pipe is not high enough to cover sensors' electrodes, the described procedure may falsely detect an anomaly, since a sudden drop in EC was detected.

During our experiments, it was noticed from the physical characteristic of the sought substances that alarms caused by conductivity change may occur only when the change was positive (rising conductivity), while drops below $500\frac{\mu S}{cm}$ indicate that sensors' electrodes are not covered fully with water. In order to cope with such situations, every EC sample is used for determining the presence of water. If the EC value drops below $500\frac{\mu S}{cm}$, every running instance of Kalman filter used in data processing is stopped. The kalman filter is resumed when a raise in the EC signal value is registered. In Algorithm 1 these thresholds are generalized as H_1 and L_1.

The running time of the Kalman filter is proportional to the size of the matrix, since its most complex operation involves inverting the matrix A. Therefore, the running time and memory footprint is constant for every provided input measurement. In practice, the running time of the anomaly detection phase, including baseline calculations and no water detection (see next subsection), is below 1 millisecond per sample when executed in the Micromole main controller.

D. SOURCE DETERMINATION ALGORITHM

As mentioned in Section V-A, the reference and master devices communicate to the master device whenever it has detected an abrupt change in the pH or conductivity of the wastewater. The master device is in charge of identifying situations where the reference device does not communicate an anomaly, but the master device has detected it. In such situations, the master device triggers the collection of a physical wastewater sample at the third device, as mentioned in Section IV.

There are two aspects that should be taken into account for proper functionality of the proposed IoT system.

1) SPECIFIC CHARACTERISTICS OF THE TARGETED PATTERN SIGNALS FOR DETECTION

Depending on the specific application, the end-user may be interested in detecting only events when pH and conductivity values abruptly changes in one specific direction. For instance, a discharge of a large volume of salt in the sewage system may cause an abrupt change in the signals given by the conductivity sensor, but detecting large spills of salt may not be of interest for the end-user.

2) PROPAGATION TIME OF THE WASTEWATER FLOW

In case the harmful wastewater is dumped before the reference device, the propagation time of the wastewater flow causes that the reference device detects an anomaly before the master device notices the abrupt change in the signal. Therefore, the master device should store the information about the anomaly from the reference device in a time-window - named B - for further analysis and correlation.

The pseudo-code of this algorithm can be seen in Algorithm 2.

Algorithm 2 Real-Time Source Determination Algorithm

Input: An alarm notification a from the IoT device i
Output: notification n

```
if i == master device then
    if buffer B does not contain a notification from the reference device then
        Remove all notifications buffer B
        return My localization to end-user
    end if
end if
Remove from B notifications exceeding an age T
Include notification a in buffer B
```

E. CASCADE MONITORING - DENSE NETWORKS

The IoT architecture explained above can be easily extended to an IoT system consisting of multiple IoT devices, allowing the monitoring of a set of consecutive neighbouring inlets in a sewer networks. In order to achieve this, the role of an IoT device in a parrticular solution depends on its position with respect to the monitored inlet. A single IoT device can take the three roles - reference, master, and sampler - for three different inlets.

As an example, in Fig. 9 we show a diagram of the functionality of Micromole IoT devices. For an inlet located

between Devices A and B, devices A, B and C are the reference, master and sampler devices, respectively. However, for an inlet located between Devices B and C, devices B, C and D are the reference, master and sampler devices, respectively.

In such scenarios, the source determination algorithm of a master devices for an inlet passes a request for a sampling action to the sampler devices. The sampler device - who may be the master device for another inlet - can already determine from such a request that the source of the wastewater flow is not adjacent to it, but further upstream. In such a way, it is possible to monitor and determine the source of pollution in a sewer mainline with N consecutive inlets, using only $N + 1$ Micromole IoT devices.

F. A NEW IoT ARCHITECTURE PATTERN

The communication restrictions in underground sewer pipes - as discussed in Section IV-C - and the time-critical requirements needed for prompt action of the actuators in our solution (sampling module) - as discussed in Section V-A - are the main aspects that motivated us for using the IoT software architecture presented above. We have revised the list of IoT architecture and design patters for common IoT problems provided by Washizaki *et al.* in [47]. To the best of our knowledge, none of the 143 IoT architecture patterns and IoT design patters cited in [47] matches the characteristics of the cascade monitoring for anomaly localization, as presented in this section. We present below a generalization of the IoT architecture pattern presented in this section, following the nomenclature of Washizaki *et al.*

- *Pattern Name:* Hop-by-hop Anomaly Detection and Actuation (HADA)
- *Intent:* Trigger an actuator as a real-time response to a detected anomaly in a sensor and actuator network
- *Context:* There is a juxtaposition between the propagation of the observed phenomena and the hop-by-hop communication link of the IoT sensor and actuator devices, in such a way that for every pair of devices with direct link communications, the phenomena can be observed sequentially across the network and the anomaly can be detected mostly by using data collected by local sensors
- *Problem:* Anomaly detection at the Edge or Cloud is costly due to difficult communication conditions for IoT sensing and actuating devices, or the latency of the communication link with the Edge or Cloud is unacceptably large or unreliable. The IoT system demands a real-time signaling of IoT actuating devices as an immediate response to a detected anomaly. Performance of the anomaly detection process and reliability of the actuator signal are of utmost importance.
- *Solution:* An anomaly detection algorithm is implemented at each IoT device using local sensing data. The outcome of the anomaly detection algorithm is communicated to the next IoT device in direction of the observed phenomena. Real-time triggering of the actuating device occurs upon detection of an anomaly that has not been detected by neighbouring nodes. Cloud services are notified about the location of the anomaly by the corresponding IoT actuating device
- *Consequences:* As measurements are not exchanged between IoT devices, the amount of energy used for data communications is low. IoT devices are capable of triggering actuating components in real-time.
- *Related Patterns:* Cloud-on-the-loop [48], Device-to-Device [48]

We consider that the HADA IoT architecture pattern can be reused for the design of IoT systems in underwater or underground environments, or in IoT applications for environmental monitoring where most of the deployed IoT devices do not have a direct connection with the Internet or the connection is unreliable, slow or costly.

G. HADA PERFORMANCE MODEL

Let us assume a dense cascade Micromole IoT network of N devices, as described in the two previous subsections, where each device requires α units of energy for sending a frame and β units of time for processing an incoming application message from the radio. Moreover, let γ be the probability of successfully transmitting a frame.

Table 6 provides models of the transmission delay, probability of losing a frame and average consumed energy in the network per detection event for the HADA architecture and compares it to the traditional Edge-based and Cloud-based architectures. The parameters ε and η represent the delay and energy used for mobile network communications, respectively.

As it can be seen, in an environment where IoT devices have very limited connectivity with the Internet, HADA performs better that Cloud and Edge-based solutions. Unfortunately, processing in the IoT device does not provide the strong computational capabilities of signal processing as offered by Cloud and Edge-based solutions.

H. MULTI-HUB MONITORING - SPARSE NETWORKS

A set of Micromole IoT devices connected to a single gateway conform a *cluster*. Two or more clusters can exchange alarms information using a dedicated Cloud or Fog service [44] (see Fig. 9), as it is performed in the EU project H2020 SYSTEM [2].

This Cloud or Fog service can relay information from the source determination algorithm of the last device of a cluster to other clusters located further downstream in the wastewater flow. In such a way, downstream clusters can determine whether a source of pollution is located at some point between the last IoT device of the first cluster and the first IoT device of the next cluster. Such information could aid investigators using the proposed solution in a more efficient deployment in the future.

For the correct operation of the Micromole system, the communication time between the last IoT device of the upstream cluster and the first IoT device of the downstream cluster must be smaller than the wastewater propagation

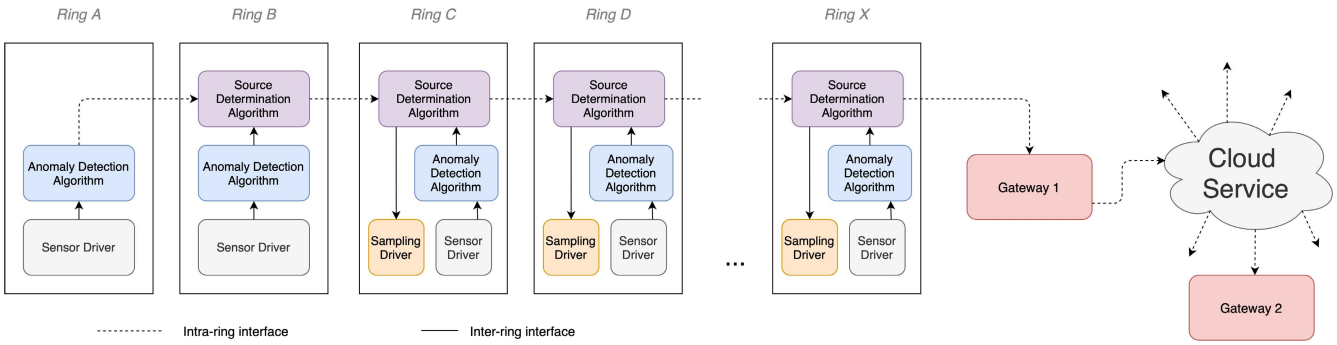


FIGURE 9. Software architecture of the distributed anomaly localization algorithm of Micromole for wastewater networks with multiple IoT devices and clusters.

TABLE 6. Performance model for HADA vs. Cloud-based approaches.

Solution	Max. transmission delay	Probability of losing a frame	Avg. Energy consumed in the network
Device-based (HADA)	2β	$1-\gamma$	2α
Edge-based	$2(N-1)\beta$	$1-\gamma^{N-1}$	$2(N-1)\alpha$
Cloud-based	$2((N-1)\beta+\varepsilon)$	$1-\gamma^{N-1}$	$2(N-1)\alpha+\eta$

TABLE 7. Substances used in simulations.

Short	Substance name	pH	EC
		-	mS/cm
H_2SO_4	Sulfuric Acid	1	1400
NaOH	Sodium Hydroxide	12	1
Na_2SO_4	Sodium Sulfate	9	12

time between these two locations. This communication time should consider the delay of the used mobile network and of the processing in the Cloud or Fog.

VI. SIMULATION RESULTS

In this section, we evaluate the capabilities of the anomaly detection algorithm of the proposed IoT system for localization in a sewage networks.

All flow and discharge simulations were performed using the software package ++SYSTEM Isar [49], which capabilities were extended by a reaction and transport model based on the concept of total alkalinity in the course of the Micromole project [1]. The anomaly localization algorithm was replicated as an R script, in order to process the flow and discharge simulation output.

A. SUBSTANCES USED IN SIMULATED DISCHARGES

We simulated discharges with three types of industrial waste substances in the real sewer network shown in Fig. 2b in order to evaluate the feasibility of the algorithm used to trigger the actuator of the IoT sampling device. The characteristics of the three substances are presented in Table 10.

B. HYDRAULIC CONDITIONS LIMITING THE EFFICIENCY OF THE ANOMALY LOCALIZATION ALGORITHM

The structure of the simulated network is shown in Fig. 2b. The network consists of 42 buildings and 15 manholes, where manhole number 16 is the sink.

In order to describe the hydraulic situation in this network immediately before the discharge event, Table 8 presents starting and ending manhole, the flow rate, water level and velocity for all sewers along the flow path for a discharge in a building connected to manhole number 09 of the network. The flow of wastewater was calculated based on the number of inhabitants for all buildings connected to each sewer and their freshwater consumption. The latter shows a typical daily pattern as can be seen in Fig. 10. Using the flow of wastewater at 03h00m of a day and the flow at 08h00m, we defined a low and a high flow scenario. This allows us to take a best-case and a worst-case view of transport and dilution processes.

C. THRESHOLDS SETUP

In order to setup the threshold of the anomaly localization algorithm, we considered the scenario when there is *no* discharge in the network. From this scenario, we obtained one time-series of EC and pH measurements for every manhole. Then, we process each time-series using the anomaly localization algorithm in R, varying the relative rising thresholds from $1500\frac{\mu S}{cm}$ to $300\frac{\mu S}{cm}$ for EC alarms, and from 1.5 to 0.3 for pH. Each time the falling threshold were set at 80% of the rising threshold, for hysteresis control. The anomaly thresholds were set at the same value as the rising thresholds. We verified that *no* simulated IoT device at any manhole yields a positive detection at any threshold.

For the remaining simulations, which will be described below, we set the rising threshold of alarms to $500\frac{\mu S}{cm}$ for EC and 0.5 for pH.

D. DISCHARGE SIMULATION SETUP

We simulated the dilution behaviour of the sewage network shown in Fig. 2b over time when industrial waste substances

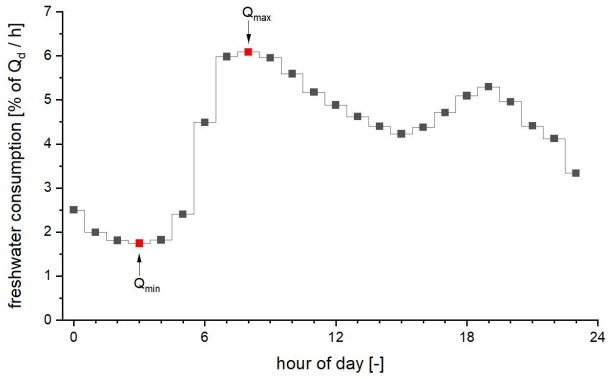


FIGURE 10. Diurnal pattern of freshwater consumption.

TABLE 8. Hydraulic characteristics of the modelled sewer network.

sewer	from	to	length m	water level cm	flow rate l/s	velocity m/s
1	9	2	58	0.4	0.02	0.09
2	2	10	38	0.4	0.03	0.09
3	10	15	27	0.4	0.03	0.09
4	15	14	6	0.5	0.05	0.13
5	14	17	29	0.5	0.07	0.13
6	17	21	42	0.5	0.08	0.13
7	21	20	36	0.8	0.16	0.16
8	20	16	38	1.2	0.16	0.20

is discharged from each building separately. A simulated event consists on the discharge of either 50L or 100L of a substance from Table 10 in a selected building at either low (3am) or high (8am) wastewater flow conditions. Therefore, there are 504 different simulation scenarios, when considering all (42) buildings, (2) flow conditions, (3) substances and (2) different discharged volumes.

From each simulated event, we obtained a set of time series of EC and pH measurements as observed through the traversing manholes in the network. An experiment is defined as a time-series of measurements of EC and pH at a single manhole from a simulation. Since there are 285 combinations of buildings and manholes, due to the flow directions of the wastewater network, the total number of experiments was 3420.

In addition to dispersion and dilution, chemical reactions between the domestic sewage and the discharged waste occur that influence the pH. The most important reactions are the neutralisation of H_3O^+ ions (R1) and OH^- ions (R2) by the HCO_3^- ions. The latter are present in the domestic sewage in a concentration of about 6 $mmol/l$.

$$HCO_3^- + H_3O+ \rightleftharpoons CO_2 + 2H_2O \quad \text{(R1)}$$

$$HCO_3^- + OH^- \rightleftharpoons CO_3^{2-} + H_2O \quad \text{(R2)}$$

Then, we processed the pH and EC time-series provided by each experiment assuming that a Micromole IoT device is installed in the corresponding manhole.

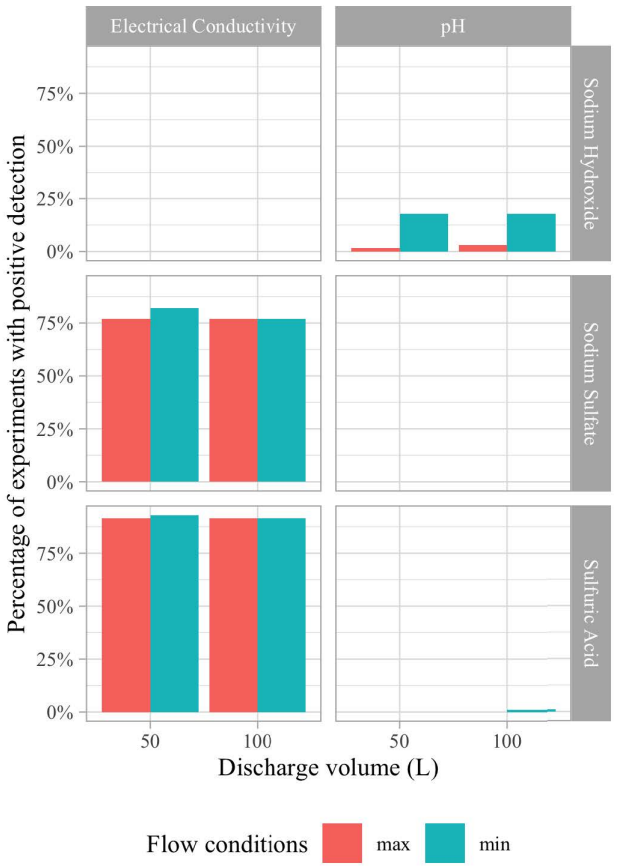


FIGURE 11. Percentage of detected experiments per discharged volume, substance, wastewater flow conditions and sensor used to detect the anomaly.

E. DETECTION RESULTS BY SUBSTANCE AND SENSOR

All simulated discharges were detected by at least one IoT device, assuming that all manholes are equipped with an IoT device. We proceed to analyze the likelihood of detecting an event, assuming that there is only one IoT device installed in the entire sewage network.

The percentage of simulated discharge experiments with positive detection for all simulations can be seen in Fig. 11. On average, 97% of the experiments corresponding to 50L or 100L of discharges of Sulfuric Acid resulted in positive detection using the EC sensor. Similarly, 82% of the experiments corresponding to 50L or 100L of discharges of Sodium Sulfate had positive detection results using the EC sensor.

Unfortunately, only 17% and 3% of the experiments of 100L of Sodium Hydroxide had positive detection outcomes using the pH sensor, if the wastewater flow conditions are low and high, respectively. We observed that only 20% of the IoT devices placed within 80 meters of the source are able to detect a change in pH levels for Sodium Hydroxide. No detection of changes in pH levels at larger distances was possible.

Even though the pH level of Sulfuric Acid and Sodium Sulfate are significantly different from the pH

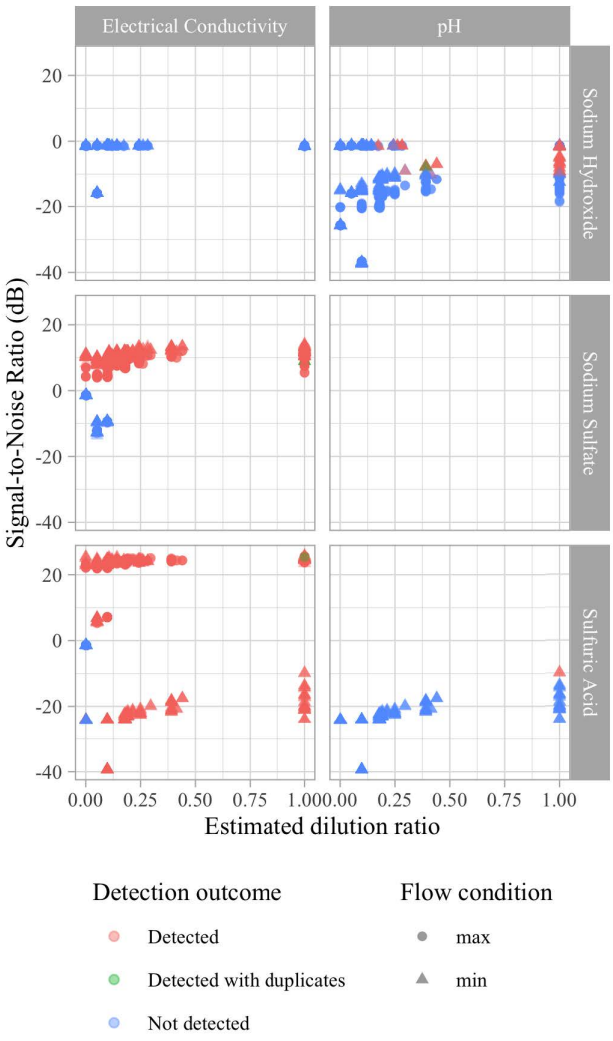


FIGURE 12. Signal-to-Noise Ratio vs. Estimated discharge dilution ratio.

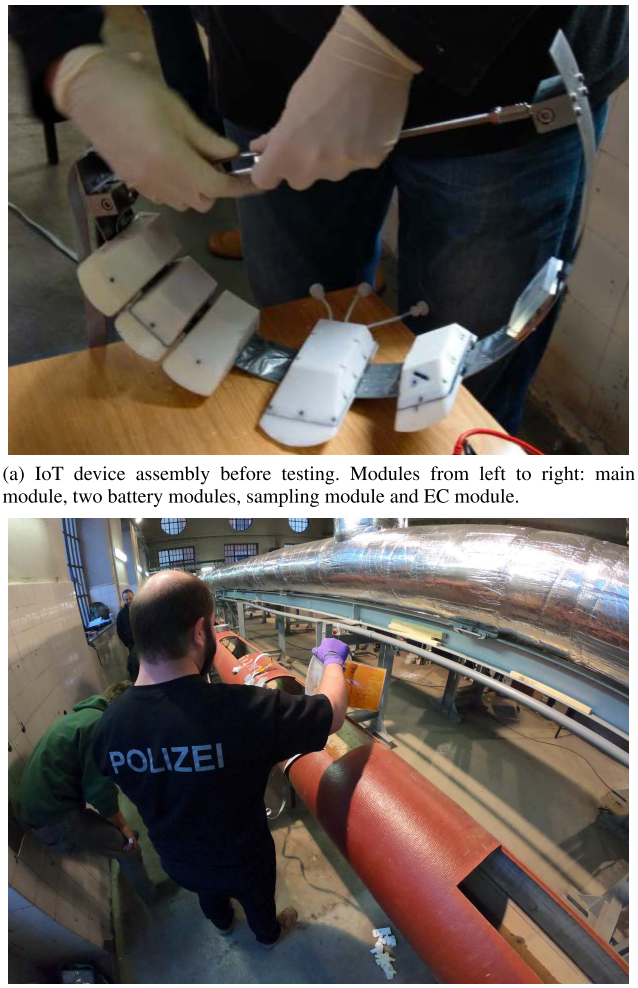
of wastewater, the proposed IoT system was able to detect only 3% of the discharges of Sulfuric Acid and none of Sodium Sulfate.

F. DILUTION AND MICROMOLE LIMITATIONS

In order to understand better the limitations mentioned in the previous subsection, we analyze the detection capabilities of the Micromole IoT devices based on the signal-to-noise ratio (SNR) and on the dilution ratio of the discharged volume at the monitoring pipe.

First, we calculated the SNR for each experiment using the method presented by Du *et al.* in [50] for Mass Spectrometry qualitative analysis: using the Mexican-hat Continuous Wavelet Transform to discover the peak amplitude and width, and the noise of the overall signal.

Then, we estimated the dilution ratio for each discharge at each sewer as follows. For this, we consider again the generated experiments with *no* discharge event, as previously



(a) IoT device assembly before testing. Modules from left to right: main module, two battery modules, sampling module and EC module.

(b) Dumping of harsh industrial waste with different pH and conductivity characteristics to the wastewater stream by a German Law Enforcement Agent.

FIGURE 13. Micromole IoT device assembly, testing and validation at KWB in December 2018.

explained in Section VI-C. Let f_e be the total wastewater volume (in liters) traversing sewer e during the simulated event, and $\phi_{i,e}$ the fraction of an unitary flow from source i going through sewer e. The dilution ratio of a discharge of volume V (in liters) from source i at sewer e, namely $z_{i,e}(V)$, was estimated as follows:

$$z_{i,e}(V) = \frac{\phi_{i,e} \cdot V + f_e}{f_e}. \tag{14}$$

The SNR in decibels and dilution ratio for each experiment can be seen in Fig. 12. As expected, we observe a correlation with the SNR and the dilution ratio, with an SNR detection threshold around 0 dB for EC and -10 dB for pH. However, we also notice no correlation between the distance and the detection outcome in most cases.

VII. SYSTEM VERIFICATION IN A REAL SEWAGE ENVIRONMENT

In this section, we describe the validation process of a working Micromole IoT system with potential end-users.

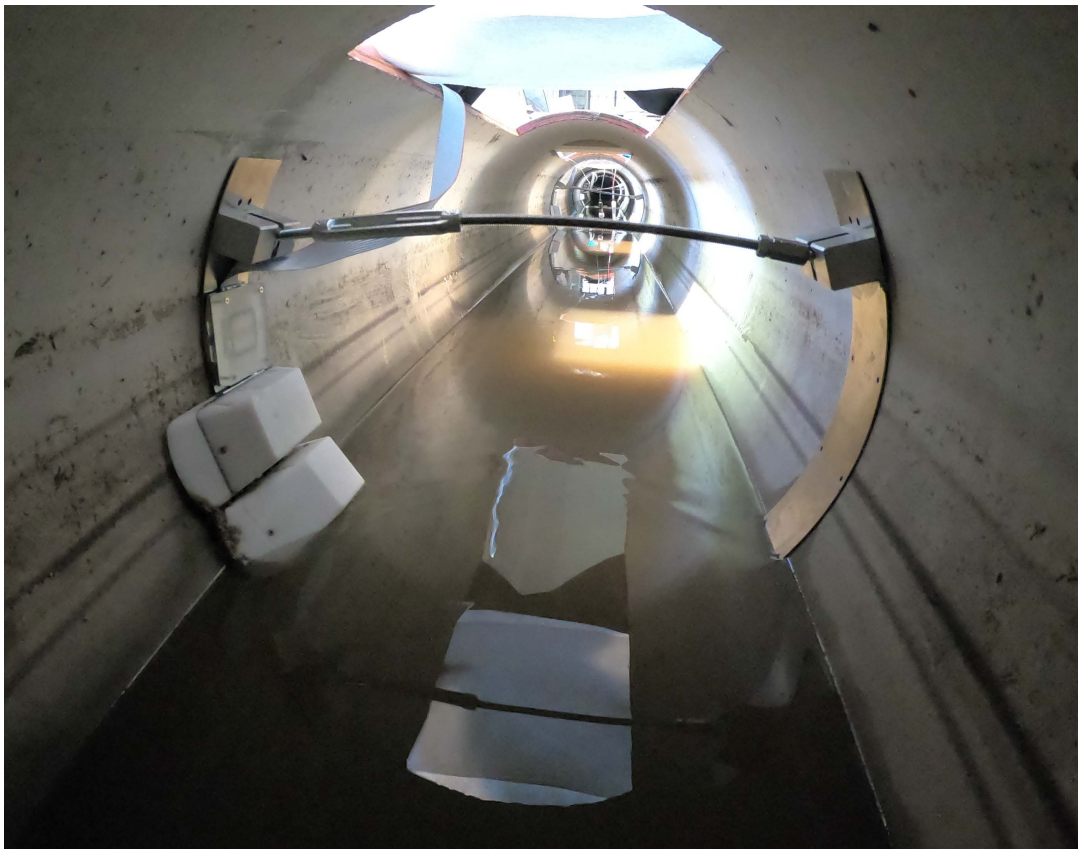


FIGURE 14. Micromole IoT device with pH and conductivity sensors measuring sewage wastewater physical parameters at KWB in December 2018. Since the Micromole IoT devices do not transmit measurements using the radio (for energy efficiency), a ribbon cable was connected to the main module of the IoT device for verification of the measurements and the experiments results through a console. The ribbon cable is not needed for operational environments, as it can be seen in Fig. 5b, since the Micromole IoT devices are completely wireless.

In order to validate the compliance with end-user requirements of the proposed IoT system with real wastewater, harsh substances and under different hydraulic conditions, numerous experiments were done at a test site at the Kompetenzzentrum Wasser Berlin (KWB) between May 2016 and December 2018. All of these experiments were guided and supervised by German and Polish Law Enforcement Agents (LEA) as potential end-users. This section describes the latest experimental setup, hydraulic conditions, discharge scenarios and results.

A. SYSTEM CONFIGURATION

The test bed at KWB consists of a gravity sewer with a nominal diameter of 350 mm and a length of 22 m. The sewer is fed with wastewater from a pumping station through a chopper pump. The wastewater is a mixture of industrial and domestic wastewater drained in combination of a separate and combined sewers. A regular check of the wastewater composition confirmed that no significant amounts of rainwater or groundwater had penetrated the connected sewers during the tests. Main indicating parameters such as Chemical Oxygen Demand (COD), pH and Electrical Conductivity exhibited quite constant values throughout all experiments. Electrical Conductivity and pH level of the wastewater in normal conditions (no discharge of harsh industrial wastewater) oscillated around $1.4\frac{mS}{cm}$ and 8, respectively.

The Master device was placed approximately 2.5 m downstream from the Reference device. The Sampler device was located 7 meters downstream from the Reference device. The point of discharge was approximately 1 m upstream from the Reference device, or at mid distance between the Reference and Master device, depending on the experiment. The discharge experiments were performed using a beaker filled with the target solution.

B. HYDRAULIC CONDITIONS

During the tests, water level, flow rate and velocity could be adjusted, in order to simulate hydraulic conditions that are typical for municipal sewer systems. By adjusting flow rate and velocity, dilution of discharged substances and flow time between our IoT devices could be varied by LEAs. Flow rate and velocity of the sewage were adjusted by weir discs and by regulating the frequency of the pump according to the values given in Table 9.

TABLE 9. Definition of hydraulic scenarios for KWB experiments.

Scenario No.	flow rate l/s	water level cm	velocity m/s
Scenario 1	8.1	11.4	0.30
Scenario 2	2.9	7.4	0.20
Scenario 3	1.0	5.1	0.12

TABLE 10. Substances used for experiments in KWB experiments.

Short	Substance	Legality	pH -	EC mS/cm
PC	Pipe Cleaner	Legal	12	22-26
IW1	Industrial Waste 1	Illegal	0	750
IW2	Industrial Waste 2	Illegal	3.6	194
IW3	Industrial Waste 3	Illegal	12	180
DW	Dish Washer effluent 3	Legal	10.2	2.9
$C_6H_8O_7$	Citric Acid	Legal	2.7	4.2
NaCl	Sodium Chloride	Legal	6.8	23

C. SUBSTANCES USED FOR SYSTEM VERIFICATION

The substances used for our experiments were chosen so that pH and electric conductivity could be influenced independently of one another. Therefore, any combination of rising and falling pH could be achieved with or without a simultaneous increase in conductivity as well as an increase in electric conductivity without a significant change in pH.

In total, LEAs chose seven different substances for our experiments, three of which could be disposed by industrial organizations and cause harm to WWTP. These substances are listed in Table 10. The substances include acid and alkaline solutions that mimic pH and electric conductivity of typical harmful industrial wastewater of different origin (IW1, IW2 and IW3 in the table) including those mentioned in I. All other substances in the table were selected in order to simulate events that could be caused by legal activities such as using a washing machine, regenerating an ion exchange resin for water softening or by cleaning of pipes with caustic or acidic agents.

A total of 14 different discharge experiments were carried out during the 5th and 6th of December 2018 with different substances, points of discharge, and flow rates. In Table 11, a description of all the executed discharge experiments is presented together with the outcome of the experiment. Almost all of the experiments consisted of a discharge of 2 Liters in a period of 15-17 seconds.

D. RESULTS ANALYSIS

In Fig. 15, the pH, EC and baselines of the reference and master devices for discharge number eight of Table 11 can be observed. In this case, both devices detect the anomaly - since the discharge was made in front of the reference device - with a time difference of nearly 2 minutes - due to the low flow rate - and with different magnitudes - due to the dilution effects previously explained. Since both devices detected the event, the master device did not request the sampling action from the sampling device.

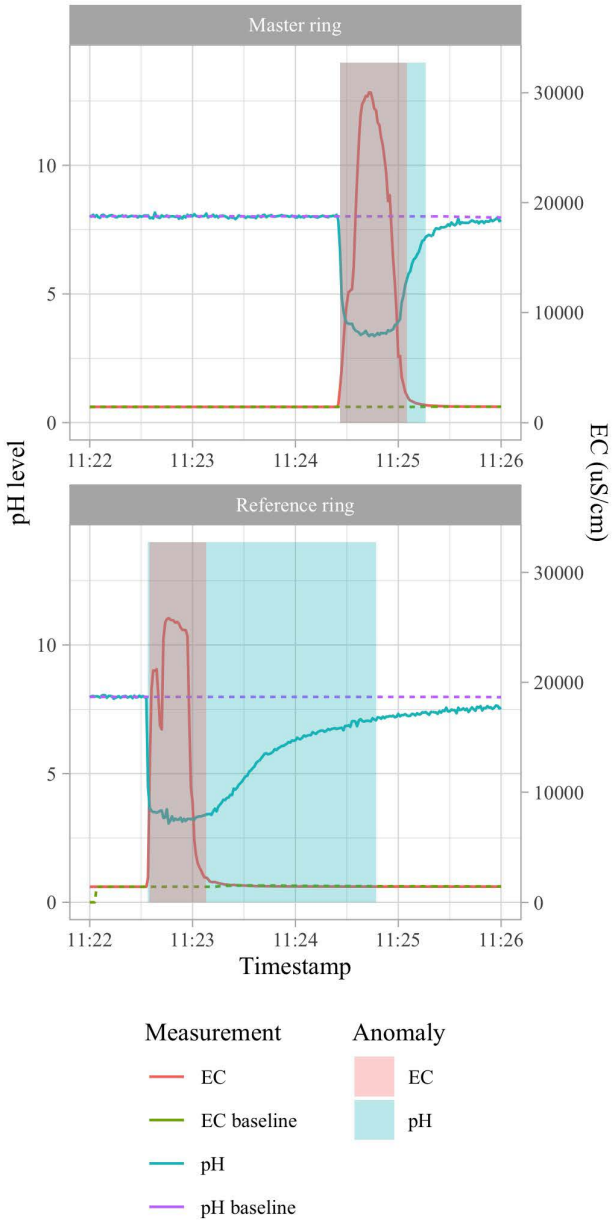


FIGURE 15. Measurements, baselines, and anomaly detection events of the Micromole Reference and Master IoT devices in KWB on 2018-12-06 following discharge number eight from Table 11.

The same outcome was achieved by discharges number 7, 12 and 14.

Discharges number 5, 6, 11, 12 and 13 - with Sodium Chloride, dish washer and citric acid - did affect the pH or EC measurements, however such measurements did not cross the relative moving threshold set for detection. Therefore, even though the substances were discharged closed and in large quantities near the master device, no detection alarm and sampling action was triggered.

For the discharges involving pipe cleaner, industrial waste 1, 2 or 3 - i.e., discharges 1, 2, 3, 4, 9 and 10 - generated

TABLE 11. Result of the discharge experiments carried out on 5th and 6th of December 2018 in KWB.

Nr.	Substance	Volume	Flow Rate	Point of discharge	pH change[a]	EC change	Outcome
1	PC	2	8.1	Between Ref. and Master devices	↑	↑	Positive
2	PC	2	8.1	Between Ref. and Master devices	↑	↑	Positive
3	IW3	2	8.1	Between Ref. and Master devices	↑	↑	Positive
4	PC	2	2.9	Between Ref. and Master devices	↑	↑	Positive
5	DW	2	2.9	Between Ref. and Master devices	↑	-	Negative
6	DW	4	2.9	Between Ref. and Master devices	↑	-	Negative
7	IW3	2	2.9	Before Ref. device	↑	↑	Negative
8	IW2	2	1	Before Ref. device	↓	↑	Negative
9	IW3	2	1	Between Ref. and Master devices	↑	↑	Positive
10	IW1	2	1	Between Ref. and Master devices	↓	↑	Positive
11	NaCl	2	1	Between Ref. and Master devices	-	↑	Negative
12	DW	2	1	Before Ref. device	↑	-	Negative
13	$C_6H_8O_7$	2	1	Between Ref. and Master devices	↓	-	Negative
14	PC	2	1	Before Ref. device	↑	↑	Negative

[a]As detected by the master device

TABLE 12. Abbreviations used in this article.

Abbreviation	Definition
6LoWPAN	IPv6 over low-power and lossy networks
ANN	Artificial neural networks
$C_6H_8O_7$	Citric acid
CoAP	Contrained application protocol
COD	Chemical oxygen demand
CPS	Cyber-physical system
DW	Dish washer
EC	Electrical conductivity
EU	European Union
F	Flow
H_2SO_4	Sulfuric acid
HADA	Hop-by-hop anomaly detection
IoT	Internet of Things
IW	Industrial wastewater
KWB	Kompetenzewasser Berlin
GC-MS	Gas chromatography with mass spectrometry
LC-MS-MS	Liquid chromatography with tandem mass spectrometry
LEA	Law enforcement agency
LoRa	Long range communications
NaOH	Sodium hydroxide
Na_2SO_4	Sodium sulfate
ORP	Oxidation redox potential
PC	Pipe cleaner
pH	Potential of hydrogen
RH	Relative Humidity
SNR	Signal-to-noise ratio
SVM	Support-vector machine
T	Temperature
TDS	Total Disolved Solids
TRL	Technology readiness level
TSS	Total Suspended Solids
Tu	Turbidity
VOC	Volatile Organic Compounds
WDS	Water distribution system
WL	Water Level
WSAN	Wireless sensor and actuator network
WWTP	Wastewater treatment plant

a detection alarm when the discharge was made between the Reference and Master device. Unfortunately, the difference in pH and EC caused by either pipe cleaner or industrial waste 1 of our experiments are similar. In both cases, the both pH and EC rises above the thresholds. In all these cases, the sampler device received a request to collect a sample from the sewage successfully.

The thresholds for EC could be adjusted in order to filter out pipe cleaner discharges as anomaly events in future experiments.

The potential end-users were asked about their level of satisfaction with the evaluated system. The end-users noted that the proposed IoT system is capable of correctly identifying harsh disposed substances in sewers and promptly triggering the sampling device, as desired.

VIII. CONCLUSION

In this article we have presented the Micromole IoT system and the distributed anomaly detection algorithm for localization of harmful wastewater discharges in the sewage. The anomaly localization algorithm has low running time and memory footprint, and it is highly robust against the natural fluctuations of pH and EC wastewater characteristics.

The unstable Internet connection in underground pipes, the lack of power grid energy supply in sewage networks, and the end-user's need for triggering the IoT device actuators in real-time brought special requirements for the design of our anomaly detection and localization algorithm. Such requirements were fulfilled by using a new IoT architecture. The new IoT architecture was generalized as an IoT architecture pattern [47], with the name Hop-by-hop Anomaly Detection and Actuation (HADA), so it could be reused in other contexts and applications.

The proposed IoT system was verified through simulations in a large sewage network and also implemented. In addition, potential end-users of the IoT system validated the system in sewer pipes with discharges of harsh industrial waste and real wastewater in KWB, Germany, several times between May 2016 and February 2019. In December 2018, a total of 14 discharges were made during the validation campaign with seven different substances at a short distance, including common domestic waste and illegal industrial waste. In all cases, the proposed IoT system was able to successfully detect and localize the source of harmful wastewater, promptly triggering the actuating device and notifying the end-user of the alarm event.

In the future work, we plan to evaluate the system in scenarios providing a different behaviour of the wastewater parameters baseline, such as: melting snow in winter, intrusion of rainwater at different levels. Moreover, we plan to evaluate the effect of heterogeneous discharge events, where

the EC and pH vary over its volume. In order to provide a better indication of illegal discharge events, we aim at considering the correlation over time of pH, EC and flow signals as a fingerprint of providing an indication of the type of substance in wastewater.

ACKNOWLEDGMENT

The authors would like to thank Krzysztof Pachowicz, Dawid Sabat, Eric Leverenz, Stefan Straube, Juan Gallardo, Frank Hauser, Thorsten Rossler, and Tim Metzner for their contribution and effort during all the testbed sessions executed in KWB.

REFERENCES

[1] H2020 Micromole Consortium. (Oct. 17, 2019). *Micromole Sewage Monitoring System for Tracking Synthetic Drug Laboratories*. [Online]. Available: http://www.micromole.eu

[2] H2020 SYSTEM Consortium. (Oct. 22, 2022). *H2020 System Synergy of Integrated Sensors and Technologies for Urban Secured Environment*. [Online]. Available: https://cordis.europa.eu/project/rcn/220304/factsheet/en

[3] F. M. Hauser, J. W. Hulshof, T. Rößler, R. Zimmermann, and M. Pütz, "Characterisation of aqueous waste produced during the clandestine production of amphetamine following the leuckart route utilising solid-phase extraction gas chromatography-mass spectrometry and capillary electrophoresis with contactless conductivity dete," *Drug Test. Anal.*, vol. 10, no. 9, pp. 1368–1382, Sep. 2018.

[4] J. G. Cormier, "Alkaline cleaning," in *Surface Engineering*, C. M. Cotell, J. A. Sprague, and F. A. Smidt, Eds. Novelty, OH, USA: ASM International, 1994, pp. 18–20.

[5] K. Ghasemipanah, "Treatment of ion-exchange resins regeneration wastewater using reverse osmosis method for reuse," *Desalination Water Treatment*, vol. 51, nos. 25–27, pp. 5179–5183, Jul. 2013.

[6] D. Butler and J. W. Davies, *Urban Drainage*, 2nd ed. London, U.K.: CRC Press, 2004.

[7] M. Van Genuchten and W. Alves, *Analytical Solutions of One Dimensional Convective Dispersive Solute Transport Equations*, vol. 1661. Beltsville, MD, USA: United States Department of Agriculture, Jun. 1982.

[8] P. Dennis, "Longitudinal dispersion due to surcharged manholes," Ph.D. dissertation, Dept. Civil Struct. Eng., Univ. Sheffield, Sheffield, U.K., 2000.

[9] J. Rieckermann, M. Neumann, C. Ort, J. L. Huisman, and W. Gujer, "Dispersion coefficients of sewers from tracer experiments," *Water Sci. Technol.*, vol. 52, no. 5, pp. 123–133, 2005.

[10] G. Currell, *Analytical Instrumentation*. Hoboken, NJ, USA: Wiley, Apr. 2008.

[11] J. P. Foley, "Equations for chromatographic peak modeling and calculation of peak area," *Anal. Chem.*, vol. 59, no. 15, pp. 1984–1987, Aug. 1987.

[12] Partech Instruments. (Nov. 19, 2020). *PH & Temperature Sensor for Water Watertechw^2 pH8000*. [Online]. Available: https://www.partech.co.U.K./product/watertechw2-ph8000/

[13] Armstrong Monitoring. (Nov. 19, 2020). *AMC-1400 Four Channel Gas Monitor*. [Online]. Available: https://www.armstrongmonitoring.com/

[14] ORI GmbH. (Nov. 19, 2020). *Mobile Sampling Basic EX 1 Mobil*. [Online]. Available: https://www.origmbh.de/en/sampling/mobile-sampling/basic-ex-1-mobile.ht%ml

[15] O. A. C. Hoes, R. P. S. Schilperoort, W. M. J. Luxemburg, F. H. L. R. Clemens, and N. C. van de Giesen, "Locating illicit connections in storm water sewers using fiber-optic distributed temperature sensing," *Water Res.*, vol. 43, no. 20, pp. 5187–5197, 2009.

[16] M. Lepot, K. F. Makris, and F. H. L. R. Clemens, "Detection and quantification of lateral, illicit connections and infiltration in sewers with infra-red camera: Conclusions after a wide experimental plan," *Water Res.*, vol. 122, pp. 678–691, Oct. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/ S0043135417305067

[17] S. De Vito, G. Fattoruso, E. Esposito, M. Salvato, A. Agresta, M. Panico, A. Leopardi, F. Formisano, A. Buonanno, P. D. Veneri, and G. D. Francia, "A distributed sensor network for waste water management plant protection," in *Sensors*, B. Andò, F. Baldini, C. Di Natale, G. Marrazza, and P. Siciliano, Eds. Cham, Switzerland: Springer, 2018, pp. 303–314.

[18] Y. Chen, C. M. Twigg, O. A. Sadik, and S. Tong, "A self-powered adaptive wireless sensor network for wastewater treatment plants," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PERCOM Workshops)*, Mar. 2011, pp. 356–359.

[19] M. Simic, G. M. Stojanovic, L. Manjakkal, and K. Zaraska, "Multi-sensor system for remote environmental (air and water) quality monitoring," in *Proc. 24th Telecommun. Forum (TELFOR)*, Nov. 2016, pp. 1–4.

[20] M. Allen, A. Preis, M. Iqbal, and A. J. Whittle, "Water distribution system monitoring and decision support using a wireless sensor network," in *Proc. 14th ACIS Int. Conf. Softw. Eng., Artif. Intell., Netw. Parallel/Distrib. Comput.*, Jul. 2013, pp. 641–646.

[21] D. Saetta, A. Padda, X. Li, C. Leyva, P. B. Mirchandani, D. Boscovic, and T. H. Boyer, "Water and wastewater building CPS: Creation of cyber-physical wastewater collection system centered on urine diversion," *IEEE Access*, vol. 7, pp. 182477–182488, 2019.

[22] D. Zhang, B. Heery, M. O'Neil, S. Little, N. E. O'Connor, and F. Regan, "A low-cost smart sensor network for catchment monitoring," *Sensors*, vol. 19, no. 10, p. 2278, May 2019, doi: 10.3390/s19102278.

[23] A. A. Cook, G. Misirli, and Z. Fan, "Anomaly detection for IoT time-series data: A survey," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6481–6494, Jul. 2020.

[24] P. Rekha, K. Sumathi, S. Samyuktha, A. Saranya, G. Tharunya, and R. Prabha, "Sensor based waste water monitoring for agriculture using IoT," in *Proc. 6th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2020, pp. 436–439.

[25] I. S. Herath, "Smart water buddy: IoT based intelligent domestic water management system," in *Proc. Int. Conf. Advancements Comput. (ICAC)*, Dec. 2019, pp. 380–385.

[26] M. Salvato, S. De Vito, S. Guerra, A. Buonanno, G. Fattoruso, and G. D. Francia, "An adaptive immune based anomaly detection algorithm for smart WSN deployments," in *Proc. 185th AISEM Annu. Conf.*, Feb. 2015, pp. 1–5.

[27] J. Zhang, X. Zhu, Y. Yue, and P. W. H. Wong, "A real-time anomaly detection algorithm/or water quality data using dual time-moving Windows," in *Proc. 7th Int. Conf. Innov. Comput. Technol. (INTECH)*, Aug. 2017, pp. 36–41.

[28] D. J. Hill and B. S. Minsker, "Anomaly detection in streaming environmental sensor data: A data-driven modeling approach," *Environ. Model. Softw.*, vol. 25, no. 9, pp. 1014–1022, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/ S1364815209002321

[29] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DBSCAN revisited, revisited: Why and how you should (Still) use DBSCAN," *ACM Trans. Database Syst.*, vol. 42, no. 3, pp. 1–21, Aug. 2017, doi: 10.1145/3068335.

[30] J. Inoue, Y. Yamagata, Y. Chen, C. M. Poskitt, and J. Sun, "Anomaly detection for a water treatment system using unsupervised machine learning," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 1058–1065.

[31] Z. X. Tian, P. Wang, L. Guo, and J. P. Jiang, "Anomaly detection of municipal wastewater treatment plant operation using support vector machine," in *Proc. Int. Conf. Autom. Control Artif. Intell. (ACAI)*, 2012, pp. 518–521.

[32] D. Jalal and T. Ezzedine, "Decision tree and support vector machine for anomaly detection in water distribution networks," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Jun. 2020, pp. 1320–1323.

[33] M. S. S. Garmaroodi, F. Farivar, M. S. Haghighi, M. A. Shoorehdeli, and A. Jolfaei, "Detection of anomalies in industrial IoT systems by data mining: Study of Christ osmotron water purification system," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10280–10287, Jul. 2021.

[34] A. Ayadi, O. Ghorbel, M. S. Bensaleh, A. Obeid, and M. Abid, "Outlier detection based on data reduction in WSNs for water pipeline," in *Proc. 25th Int. Conf. Softw., Telecommun. Comput. Netw. (SoftCOM)*, Sep. 2017, pp. 1–6.

[35] V. Fehst, H. C. La, T.-D. Nghiem, B. E. Mayer, P. Englert, and K.-H. Fiebig, "Automatic vs. manual feature engineering for anomaly detection of drinking-water quality," in *Proc. Genetic Evol. Comput. Conf. Companion*, Jul. 2018, pp. 5–6, doi: 10.1145/3205651.3208204.

[36] K. Joslyn and J. Lipor, "A supervised learning approach to water quality parameter prediction and fault detection," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 2511–2514.

[37] M. Sonrani, M. Abbatangelo, E. Carmona, G. Duina, M. Malgaretti, E. Comini, V. Sberveglieri, M. P. Bhandari, D. Bolpagni, and G. Sberveglieri, "Array of semiconductor nanowires gas sensor for IoT in wastewater management," in *Proc. Workshop Metrol. Ind. 4.0 (IoT)*, Apr. 2018, pp. 68–71.

[38] J. Gallardo-Gonzalez, A. Baraket, S. Boudjaoui, T. Metzner, F. Hauser, T. Rößler, S. Krause, N. Zine, A. Streklas, A. Alcácer, J. Bausells, and A. Errachid, "A fully integrated passive microfluidic lab-on-a-chip for real-time electrochemical detection of ammonium: Sewage applications," *Sci. Total Environ.*, vol. 653, pp. 1223–1230, Feb. 2019.

[39] E. Leverenz, K. Becker, M. Koch, S. Straube, H. Poetter, and K. Lang, "Energy autarkic wireless sensor node for reliable long-term exposure to domestic waste water in a sewage system," in *Proc. Sensors Measuring Syst., 19th ITG/GMA-Symp.*, Jun. 2018, pp. 1–4.

[40] RIOT. (Jan. 30, 2019). *Riot-OS*. [Online]. Available: http://https://riot-os.org/

[41] Z. Shelby, K. Hartke, and C. Bormann, *The Constrained Application Protocol (CoAP)*, document RFC 7252, Jun. 2014. [Online]. Available: https://rfc-editor.org/rfc/rfc7252.txt

[42] G. Montenegro, C. Schumacher, and N. Kushalnagar, *IPv6 Over Low-Power Wireless Personal Area Networks (6LoWPANs): Overview, Assumptions, Problem Statement, and Goals*, document RFC 4919, Aug. 2007. [Online]. Available: https://rfc-editor.org/rfc/rfc4919.txt

[43] F. Solano Donado, "On the optimal calculation of the Rice coding parameter," *Algorithms*, vol. 13, no. 8, p. 181, Jul. 2020, doi: 10.3390/a13080181.

[44] I. Martinez, A. S. Hafid, and A. Jarray, "Design, resource management, and evaluation of fog computing systems: A survey," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2494–2516, Feb. 2021.

[45] J. Botero-Valencia, L. Castano-Londono, D. Marquez-Viloria, and M. Rico-Garcia, "Data reduction in a low-cost environmental monitoring system based on LoRa for WSN," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3024–3030, Apr. 2018.

[46] M. H. Syed, E. B. Fernandez, and M. Ilyas, "A pattern for fog computing," in *Proc. 10th Travelling Conf. Pattern Lang. Programs (VikingPLoP)*, 2016, pp. 1–10, doi: 10.1145/3022636.3022649.

[47] H. Washizaki, S. Ogata, A. Hazeyama, T. Okubo, E. B. Fernandez, and N. Yoshioka, "Landscape of architecture and design patterns for IoT systems," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10091–10101, Oct. 2020.

[48] G. Bloom, B. Alsulami, E. Nwafor, and I. C. Bertolotti, "Design patterns for the industrial Internet of Things," in *Proc. 14th IEEE Int. Workshop Factory Commun. Syst. (WFCS)*, Jun. 2018, pp. 1–10.

[49] Tandler.Com. (Nov. 21, 2020). *Software for Water Management ++ Systems Isar*. [Online]. Available: https://www.tandler.com/en/page/software/plus-plus-systems

[50] P. Du, W. A. Kibbe, and S. M. Lin, "Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching," *Bioinformatics*, vol. 22, no. 17, pp. 2059–2065, 2006, doi: 10.1093/bioinformatics/btl355.

FERNANDO SOLANO received the bachelor's degree in computer science and engineering from Universidad del Norte, Colombia, in 2003, and the Ph.D. degree in information technologies from the University of Girona, Spain, in 2007.

He is currently an Assistant Professor with the Warsaw University of Technology, Warsaw, Poland, and the CEO and the Owner of Blue Technologies. He has spent eight months in research collaboration with Drexel University, USA, and Cisco System, and later, six months with IBBT, Belgium. Since 2007, he has published nearly 80 scientific papers, five journals, one patent application and a book chapter. He serves as a reviewer for several scientific journals including IEEE TRANSACTIONS ON NETWORKING, IEEE TRANSACTIONS ON COMMUNICATION, and *Journal on Computer Networks* (Elsevier). He has coordinated two European Projects: the FP7 Project Goldfish and the H2020 Project Micromole. He is also the Work-Package Leader in the H2020 Project SYSTEM. In addition, he has been a Technical Coordinator of a EUREKA-CELTIC Project Mango, besides participating in COST Actions (GRAAL) and several national Spanish projects. He has coordinated work-packages on several software development projects in Colombia for private organizations (including video conferencing applications, emails, website, and business process management), between 2001 and 2004.

STEFFEN KRAUSE received the Diploma degree in process chemistry from TH Merseburg, Germany, in 1990, and the Ph.D. and the post doctoral degrees in civil engineering from Bundeswehr University Munich, Munich, Germany, in 2001 and 2011, respectively.

He is currently an Assistant Professor and the Head of the Laboratory of the Chair of Sanitary Engineering and Waste Management with Bundeswehr University Munich. For about two decades, he worked on membrane technology for drinking water treatment with a focus on small scale supply systems. In the last years, he worked on critical infrastructure, such as sewer networks and water supply, including emergency preparedness planning. He was part of the European H2020 Project Micromole and is currently the Work-Package Leader in the H2020 Project SYSTEM for the part of sewage network modeling.

CHRISTOPH WÖLLGENS received the bachelor's degree in waste management engineering and the master's degree in waste management engineering with focus on soil protection and water management from RWTH Aachen University, Germany, in 2014 and 2019, respectively.

He is currently a Research Assistant at the Chair of Sanitary Engineering and Waste Management with Bundeswehr University Munich, Munich, Germany. His current research interests include wastewater discharge, sewer networks, and modeling. His current work focuses on sewage network modeling within the Horizon 2020 Project SYSTEM.