

Article Cost-Efficient Coverage of Wastewater Networks by IoT Monitoring Devices [†]

Arkadiusz Sikorski¹, Fernando Solano Donado² and Stanisław Kozdrowski^{1,*}

- ¹ Institute of Computer Science, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland
- ² Institute of Telecommunications, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland
- * Correspondence: stanislaw.kozdrowski@pw.edu.pl; Tel.: +48-22-234-5048
- † This paper is an extended version of our paper published in International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Hvar, Croatia, 23–25 September 2021, https://doi.org/10.23919/SoftCOM52868.2021.9559098.

Abstract: Wireless sensor networks are fundamental for technologies related to the Internet of Things. This technology has been constantly evolving in recent times. In this paper, we consider the problem of minimising the cost function of covering a sewer network. The cost function includes the acquisition and installation of electronic components such as sensors, batteries, and the devices on which these components are installed. The problem of sensor coverage in the sewer network or a part of it is presented in the form of a mixed-integer programming model. This method guarantees that we obtain an optimal solution to this problem. A model was proposed that can take into account either only partial or complete coverage of the considered sewer network. The CPLEX solver was used to solve this problem. The study was carried out for a practically relevant network under selected scenarios determined by artificial and realistic datasets.

Keywords: sewer network; wireless sensor network; internet of things; combinatorial optimization; mixed integer programming

1. Introduction

Wastewater networks are a critical infrastructure: an asset essential for the functioning of society and the economy. Its proper functioning can be impaired by several threats, such as sewage pipe leaks or ruptures, malfunctioning of the wastewater treatment plant (WWTP), etc.

One of the most important threats for its correct functioning in an urban environment relates to the illegal disposal of harsh chemicals in the sewer network. These chemicals may spread beyond the sewer network, and since the capacity of the sewage network and of the WWTP is limited, these chemicals may leak and contaminate groundwater reservoirs, or damage the wastewater treatment plants and render it offline. Examples of unlawful activities of industrial organizations in the sewage network are discharges of: (a) sulfuric acid (H₂SO₄), resulting from the etching of semiconductors, accumulator acid, or the production of organic chemical substances [1]; (b) sodium hydroxide (NaOH), resulting from cleaning of surfaces in metal processing in industrial applications [2]; (c) sodium sulfate (Na₂SO₄), resulting from regeneration of cation exchange resins, which are used for softening of water in industrial water treatment [3]. Illegal discharges of such dangerous harsh industrial waste into sewage networks could be harmful for the biological stage of WWTP, its personnel, sewer pipes, and the general public.

Detecting illegal discharges of any of three substances mentioned above can be performed by sampling the wastewater with commercial pH and Electrical Conductivity (EC) sensors. Nevertheless, due to wastewater dilution and mixing effects in sewer pipes throughout the sewage network, the concentration of such substances may be below the



Citation: Sikorski, A.; Donado, F.S.; Kozdrowski, S. Cost-Efficient Coverage of Wastewater Networks by IoT Monitoring Devices. *Sensors* 2022, 22, 6854. https://doi.org/ 10.3390/s22186854

Academic Editor: Stefano Mariani

Received: 27 July 2022 Accepted: 7 September 2022 Published: 10 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). minimum detection threshold of such sensors several hundred meters downstream in a populated sub-catchment area. Therefore, it is important to monitor the wastewater composition at multiple points in the sub-catchment area.

As a result, several portable Internet of Things (IoT) systems for monitoring wastewater composition have been proposed in recent years [4–18]. These IoT systems are adapted for working at manholes or main sewer lines, and usually comprise a set of sensors (electrochemical sensors, optical sensors, mass spectrometry, ion spectrometry, etc.) for detecting the presence or concentration of specific marker pollutants.

One of such IoT systems is the Micromole system [4,5]. The Micromole system consists of one or more battery-operated devices mounted at sewer main lines. Each device is equipped with pH and Electrical Conductivity (EC) sensors, specially designed for its operation in flowing wastewater [19]. The micromole device is composed of several detachable replaceable modules. In Figure 1 a micromole device comprising five of such modules can be observed. Some of these modules contain batteries, while others contain sensor electronics.



Figure 1. Micromole ring with five modules attached for measuring sewage wastewater physical parameters. From left to right, the attached modules are: battery module, wireless communication module, pH sensor module, Electrical Conductivity sensor module, and a Water Level sensor module.

This articles focuses on the planning of an cost-effective positioning of a network of IoT devices monitoring a sewage network. Below we provide an overview of the most recent methods proposed in the literature for the planning of monitoring devices in the sewage network.

This paper is organized as follows. Section 2 presents a description of the most relevant works on the subject. In Section 3, the problem is described and the model is presented with a brief explanation of the dispersion phenomena in wastewater networks. In Section 4 we describe a set of numerical experiments realized within a sewage network in the subcatchment area of an European city. Section 5 provides the conclusions of our findings.

2. Related Work

The SIMONA project [20] has as one of its main goals proposing methods and algorithms for the planning of water quality monitoring stations in sewer systems. Banik et al. propose a set of solutions [21–24], all of which share the following approach. First, the authors consider as input to the problem a set of time-series of measurements, where one

time-series consists of the measurements that would be observed at a given point in the sewage network if one potential source in the network makes a discharge. The measurements provide an indication of the quality of the wastewater, e.g., Electrical Conductivity, following certain given hydraulic conditions. Each measurement of the time-series is then quantized in discrete steps: rounding each measurement to its nearest value in the new scale. As a result, the number of potential different input values is constrained. Next, Banik et al. calculate the information entropy, or information content, of each time-series. After the previously described procedure for pre-processing is executed, Banik et al. consider a dual-objective optimization problem for the placement of the sensor devices. The objective function and meta-heuristic used for finding these solutions vary among Banik et al. contributions, which we summarise below.

In Ref. [21], the two objectives are: (1) maximum information content attained by a group of monitoring stations and (2) minimum the dependency among the monitoring stations. The first objective is achieved by maximizing the joint entropy of the selected monitoring stations, while the second one is attained by minimizing the total correlation of the chosen solution subset of monitoring stations. The set of Pareto optimal solutions is found by using an NSGA-II heuristic. According to Ref. [22], the final decision of selecting the set of monitoring stations from this Pareto front is made by maximizing the amount of information gained by a set of monitors, maintaining the consistency of the selected set of monitors for both variables (concentration and detection time) and having minimum total correlation within a set. The information theory approach taken by Banik et al. has been previously used in related areas [25,26].

In Refs. [22,23] Banik et al. extend their study by considering two additional objectives: detection time of an anomaly and reliability of the solution. The objective related to the detection time aims at minimizing the elapsed time from the discharge event until its detection, when using a fixed number of sensors. The objective related to reliability is related to the number of contamination scenarios that could be potentially correctly detected. Solution to this multi-objective optimization problem were found using the greedy algorithm proposed by Alfonso et al. in Ref. [27], originally designed for other applications.

Our previous work [28] presented the problem of optimising the number of IoT devices in a sewer network, while considering a fixed battery capacity for our sensors in a way that any potential illegal discharge in the sewage network could be detected. In this article we, instead, consider a partial network coverage and include the limitations imposed by sewage physical dimensions on the allocated battery capacities and sensor sampling rates. To the best of our knowledge, this article is the first one in the literature tackling such a problem.

Even though there are design methods in the literature—e.g., Genetic Algorithms [29,30], or Particle Swarm Optimization Algorithms [31]—for solving network coverage problems using Wireless sensor networks (WSN), none of them exploit the flow propagation properties and hydraulic dilution phenomena, as discussed in this article, in their solutions.

3. Related Background Knowledge and Proposed Methods

In this manuscript, we consider the problem of optimising the positioning of a wireless sensor network for monitoring the sewer network. In addition, the tackled problem also considers the appropriate allocation of the battery capacity of each sensor device, while considering energy requirements.

Two important requirements for the design of such sensor devices are: (1) to allow its placement in sewer mainline pipes of at least 250 mm of diameter without blocking the flow of sewage, and (2) ease of sensor and battery replacement. Micromole devices fulfil the first requirement by adopting a ring mechanical structure, as shown in Figure 1. Micromole devices fulfil the second requirement by housing electronics into a set of interchangeable modules, each of which share the same dimensions and electronic interconnections. These modules are mechanically and electronically interconnected through the ring mechanical structure, as shown in Figure 1. Since all modules have the same volume, the energy

capacity that can be stored using batteries is the same for each module. Nevertheless, even though the energy capacity provided by any module is the same, the number of modules that can be attached to a Micromole device varies and largely depends on the circumference of the ring and, hence, is limited by the pipe diameter where it will be installed: wider pipes allow for a placement of more battery modules for a single device.

The energy consumption of the Micromole device is mostly dependent on the sampling frequency used by its sensors. The sampling frequency shall be set as to avoid situations where the device fails to notice a short discharge, due to its proximity to the source, fast flow speed, or short discharge time. Mitigating such situations can be achieved by assuming that the sampling frequency is dependent on the sewage flow velocity: fast flowing sewage requires high sampling frequency.

In this article we consider that the overall cost of a sensor device comprises the cost of the sensor electrodes themselves—which we consider as a fix cost per sensor device unit—and the cost of the chosen number of allocated battery units.

3.1. Pollution Detection and Sensor Localisation

In this article we assume that there is only a single polluting source at a time in the monitored sewer network. This is motivated by the fact that illegal discharges or wastewater pollution is a rare event. Nevertheless, the location of the polluting source, if present, is unknown.

The concentration of an injected pollutant fluctuates from pipe to pipe and, over time, due to the dispersion and dilution effects caused by the mixing of inflows in the sub-catchment area. This effect can be observed in Figure 2, where the EC of wastewater is shown for 82 measuring points downwards a polluting source, from which 50 L of sulphuric acid were disposed.



Figure 2. EC broadening and flattening caused by dispersion as seen at different measuring points, when 50 litres of sulphuric acid are discharged 81 manholes upstream from the sink point of the network shown in Figure 3 in low wastewater flow conditions.

A similar effect can be observed when measuring the amount of the diluted pollutant at the same pipe at different points in time during the day: as social and industrial activities demand more usage of water at certain hours, the amount of total flow in a pipe increases and so does the dilution factor of the pollutant. We refer to *flow conditions* as the amount of flow on every pipe at a given point of time.



Figure 3. Sewage network used for numerical experiments.

Due to the dilution effects and limited sensitivity of the sensor devices, the pollutant can only be detected in those pipes where the diluted amount of the substance exceeds the minimum limit of detection of the sensor. We say that a sensor located at pipe *e covers* a potential pollution source s_i when considering flow conditions f, if the sensor can detect the injection of a pollutant with an anomaly detection method using its collected time-series of sensor measurements. For the purpose of this study, we use a simple threshold criteria as our anomaly detection method: if a measured value exceeds a predefined threshold Q, then the sensor can detect the injection of the pollutant. The usage of a simple threshold as

an anomaly detection method does not exclude the usage of more complex methods for anomaly detection—such as those based on pattern matching or Artificial Intelligence [32], for instance, or data fusion [33,34].

As a consequence, and given that the flows of wastewater is acyclic in a sewage network, the set of pipes where the pollution from a particular source can be detected form a directed acyclic sub-graph, $\mathcal{G}(s_i)$, of the sewage network. It shall be noted that for two polluting sources s_i and s_j , the corresponding sub-graphs, $\mathcal{G}(s_i)$ and $\mathcal{G}(s_j)$, may have edges in common. If a sensor device is installed in a common edge, it is not possible to discern whether the detected pollutant originates from either s_i or s_j , by only using our threshold criteria.

3.2. Model Description

Made assumptions in terms of domain language. These are expressed mathematically in the next sub-section.

Nodes:

- a set of nodes denoted as \mathcal{V} is defined, each node of this set represents a sewer manhole;
- a few nodes are distinguished as outlet nodes of the given sewer network;
- a set $V_s \subset V$ is defined and represents nodes that can be sources of undesirable substances. Edges:
- a set of directed edges is defined, each edge represents a pipe in the sewer network;
- any two nodes can have at most only one *direct* connection;
- edges can be marked as private or public. In Figure 4, dashed lines represent private pipes and solid lines represent public ones;
- each of these edges is characterized by a parameter that determines the size/flow capacity of water in each pipe;
- each pipe has a limited cross-area section, which limits the number of slots that can be used for attaching sensors and batteries in a single device. Sensors and batteries can only be installed on a ring device. Such a ring has a fixed cost.

Sensors:

- sensors detect undesired substances in the sewage, they are to be installed in the edges of the graph;
- only public edges are eligible for sensor installation while the private ones are not;
- each sensor has a fixed cost of installation;
- it is not known a priori how many sensors are required;
- sensors can only detect the contamination if the concentration in the pipe is not too low, since each sensor has a detection threshold. Each potential source of contamination is associated with a subgraph where the contaminant will be effectively detected and only there it makes sense to install sensors;
- discharge of undesired substances is a rare event and there can be only one at a time; there is no need to install sensors in a way that several sources can be distinguished;
- each sensor can sample the sewage at a given frequency—the bigger the flow, the more sensors will be needed to sample the flowing sewage—linearly more (one sensor is enough to sample the sewage having velocity 1 m/s but flow having the velocity 2 m/s requires two sensors).

Battery:

- sensors require batteries to run;
- the number of batteries required by each sensor depends on where the sensor will be installed;
- the number of batteries depends functionally on the sampling frequency, which depends on the flow rate and size of the pipe where the sensor will be installed;
- each battery has a fixed cost;

- size of the pipe limits the number of batteries that can be installed in the pipe as well as the number of sensors; the number of sensors and batteries combined is limited by the number of slots on the ring;
- the installed batteries deplete linearly, all at once.

Coverage of the sources:

- all potential pollution sources in the sewage network should be covered, i.e., any contamination discharge should be detectable by at least one sensor;
- one sensor can cover several sources since discharge from only one of them can happen at the time and there is no need to distinguish them;
- definition of coverage: for each source node *s* ∈ *V_s* there is defined a subgraph *G_s* where it makes sense to install sensors. If there is at least one sensor in each such subgraph, we satisfy the coverage condition;
- any solutions where any pollution source is not covered is not approvable;
- the coverage constraint is satisfied in Figure 4—the sensor covers both pollution sources that are denoted as red triangles. There is no need to put a sensor in the second leg of the network.



Figure 4. Network coverage. Assuming that $V_1 = 2, 4, 6, 7$ and $V_2 = 4, 5$, the sensor installed in pipe 4 is enough to cover both sources. Installing it in pipe 5 would cover only the second source.

Objective:

- Minimise the total cost of installing the sensors together with the cost of purchasing batteries for each sensor;
- We are interested in covering all or parts of the network, so that there is no potential source of contamination that is not detected by at least one sensor.

3.3. Mixed Integer Programming Model

The Mixed Integer Programming (MIP) method is proposed to solve the presented problem [35]. The advantage of this method is that it guarantees an optimal solution, as it searches the entire space of admissible solutions to the given problem. In general, the disadvantage of this method is that it often takes a long time to calculate the optimum [36]. In this case, in the problem under consideration, the MIP method performs quite well, even for networks with a large number of nodes.

The following will present the proposed model in mathematical terms. We will describe the definitions of the indices, sets, constants, and variables that appear in this model before the objective function and the necessary constraints are presented. We will operate with the indices e and s. The former refers to the edges and the latter to the nodes, which are the sources of pollution in the network under consideration. The sets, variables, and constants, on the other hand, are presented in Tables 1, 2 and 3, respectively.

Set	Description
\mathcal{V}	Each vertex $s \in \mathcal{V}$ of the graph \mathcal{G} represents a manhole
ε	The set of directed edges of graph $\hat{\mathcal{G}}$, which represent the sewage pipes over which the sewage flows
\mathcal{V}_{s}	The set of vertices that could be potential sources of pollution; $\mathcal{V}_s \subset \mathcal{V}$
\mathcal{E}_s	The set of edges at which the concentration of pollutants allows effective detection of harmful substances after they have been emitted from the vertex <i>s</i> , also known as proximity; $\mathcal{E}_s \subset \mathcal{E}$.

Table 1. Sets description.

 Table 2. Variables description.

Variable	Description
α _e	$\alpha_e \in \mathcal{N}$; variable indicating
	how many sensors have been installed at edge $e \in \mathcal{E}$
β_e	$\beta_e \in \mathcal{N}$; variable indicating
	how many batteries have been installed at edge $e \in \mathcal{E}$
γe	Binary; equal 1 if the ring is installed on the edge $e \in \mathcal{E}$;
	0 otherwise
δ_s	Binary; equal 1 if the source $s \in V_s$ is covered;
	0 otherwise.

 Table 3. Constants description.

Constant	Description
Λ_e	Number of slots in the ring installed on the edge $e \in \mathcal{E}$
Γ_e	The cost of installing a ring on the edge $e \in \mathcal{E}$
Α	The cost of one sensor
В	The cost of one battery
Ω_e	Total battery life at the edge $e \in \mathcal{E}$; expressed in sec;
	We assume that the sensor samples continuously;
	an example value is 10^6 s.
Φ_{e}	Sampling frequency of the sensor at the edge $e \in \mathcal{E}$;
	e.g., once per minute, then $\Phi_e = 1/60$
Θ	Capacity of one battery; expressed in the number of samples made,
	e.g., $\Theta=10^5$, assuming that the batteries are the same
	on each edge $e \in \mathcal{E}$
Π	Percentage of source coverage.

Objective:

$$\min\left\{\sum_{e\in\mathcal{E}} \left(A\cdot\alpha_e + B\cdot\beta_e\right) + \Gamma_e\sum_{e\in\mathcal{E}}\gamma_e\right\}$$
(1)

Constraints:

$$\alpha_e + \beta_e \le \Lambda_e \cdot \gamma_e \qquad \forall_{e \in \mathcal{E}} \tag{2}$$

$$\sum_{e \in \mathcal{E}_s} \alpha_e \ge o_s \qquad \forall_{s \in \mathcal{V}_s} \tag{3}$$

$$\sum_{e \in \mathcal{E}_s} \gamma_e \ge \delta_s \qquad \qquad \forall_{s \in \mathcal{V}_s} \tag{4}$$

$$\beta_{e} \cdot \Theta \ge \Omega_{e} \cdot \Phi_{e} \cdot \alpha_{e} \quad \forall_{e \in \mathcal{E}}$$

$$\sum \delta \sum [\Pi_{e} \mid \mathcal{V}_{e} \mid] \qquad (6)$$

$$\sum_{s \in \mathcal{V}_s} \delta_s \ge |\Pi \cdot |\mathcal{V}_s|| \tag{6}$$

Formula (1) represents the cost function of the presented problem, which is subject to minimisation. Constraint (2) guarantees us that the number of slots in the ring installed on edge *e* does not exceed the available number of slots. Then, constraint (3) means that each potential source is covered by at least one sensor. Constraint (4) tells us that at least one ring must be installed on each edge where the concentration allows detection of harmful substances, while constraint (5) ensures that the capacity of all batteries must be greater than the lifetime and sampling frequency of the edge *e*. Finally, constraint (6) indicates the percentage of sources to be covered.

4. Experimental Results and Discussion

The proposed mathematical model was tested with two different datasets, each of which was derived from the same sewage network, which is depicted in Figure 3. The sewage network consists of 3297 manholes, 3343 pipes, and 1315 sources of pollution.

The first dataset uses a sub-graph of the base network and consists of 1124 pipes and 402 pollution sources while the second one uses the whole network.

Sections 4.1 and 4.2 describe how \mathcal{E}_s sets were created—using discharge simulations and a simplified dispersion model respectively. Section 4.3 describes how sampling frequencies were pre-computed for both datasets. The following two subsections provide results and discussion of the actual cost optimization process using the linear model.

4.1. Dataset 1: Simulated Discharges and Dispersion Modelling

All flow and discharge simulations were performed using the software package ++SYSTEM Isar [37], which capabilities were extended by a reaction and transport model based on the concept of total alkalinity in the course of the Micromole project [4].

Due to computational constraints of the ++SYSTEM Isar system, it was not possible to simulate a discharge from every single building in the sub-catchment area. Instead, a subset of 402 buildings were chosen as potential sources of pollution. From every single potential source of pollution, we simulated discharges of 50 L of sulphuric acid, with pH 1 and EC 1400 mS/cm, with low flow conditions and with high flow conditions. Low flow conditions— f_L —represent the amount of flow found in this sewage network at 03 h 00 m, while high flow conditions— f_H —represent the amount of flow found in this sewage network at 08 h 00 m during a normal work day.

For establishing the sensor coverage for every particular pipe, we set a threshold for the EC value. In our experiments, we evaluated three different threshold values for EC: $Q_1 = 2 \text{ mS/cm}$, $Q_2 = 3 \text{ mS/cm}$, and $Q_3 = 4 \text{ mS/cm}$, where the normal EC value of wastewater is nearly 1.3 mS/cm. As a result, the combination of the two flow conditions and the three EC threshold values results in six different scenarios that we evaluate below.

4.2. Dataset 2: Simplified Dispersion Model

Since discharge simulation is a heavy computational task, an inherited method of proximity generation was introduced to provide test data for a greater number of pollution sources. The algorithm of generating \mathcal{E}_s sets is presented as Algorithm 1.

Alg	Algorithm 1 Simplified generation of proximities					
1:	1: function GENERATEPROXIMITIES(\mathcal{G}, k)					
2:	$\mathcal{V}_s \leftarrow findSourceNodes(\mathcal{G})$					
3:	for $s \in \mathcal{V}_s$ do					
4:	$d \leftarrow findNearestDrainNode(s, \mathcal{G})$					
5:	$p \leftarrow findShortestPathBetween(s, d, G)$					
6:	$\mathcal{E}_s \leftarrow takeEdgesFromPath(p,k)$					
	$\text{return } \{\mathcal{E}_s \forall s \in \mathcal{V}_s\}$					

The above pseudocode requires some commentary:

- 1. All source nodes should be found or defined at the beginning; a source node has exactly one outcoming edge and no incoming edges;
- 2. For each source node *s* the shortest path between *s* and the closest drain node *d* needs to be found. It is the shortest in the terms of lowest number of edges;
- 3. Each shortest path is shortened and only the first *k* edges are taken. We assume that *k* pipes is enough for a pollutant to become undetectable by a sensor. This simplification is precise enough since pipes in the neighbourhood of each source have comparable lengths. *k* is chosen based on simulated data. We decided to test cases for k = 10, 20, 30, 40 since the average and the median length of a path in simulations was about 20 edges.

This method does not require dispersion simulation, which is computationally challenging. Instead, it uses simple graph algorithms, such as shortest path finding. The paths are limited to a length obtained from the simulations run using the smaller network.

4.3. Determining Sampling Frequencies for Both Datasets

Sampling frequencies in each pipe had to be calculated for both datasets. The sampling frequency in pipe *e* is affected by two factors:

- 1. The volume of sewage flowing through the pipe denoted as u_e . The greater the quantity of sewage in the pipe, the greater sampling frequency needs to be;
- 2. The area of the pipe's section, denoted as Ψ_e , calculated using a standard formula for disk area. The greater the section's area, the slower the flow in the pipe, so the sampling frequency can be lower.

Assuming that each source *s* continuously adds 1 discrete flow unit of sewage to the network, the flow values are generated as follows (see Figure 5):

- 1. For each *e*: set flow value $u_e = 0$;
- 2. For each source *s*: find the shortest path between *s* and the closest drain node *d*;
- 3. For each path p: for each edge e belonging the path p, increase flow value u_e by 1 unit.



Figure 5. Flow units propagating through the network. The number over the edge is the number of flow units in the pipe. The greater number of flow units next to the outlet node means that a bigger volume of sewage flows in that part of the network when compared to pipes next to the sources.

Finally, sampling frequencies can be determined using the formula $\Phi_e = (\Phi_b + \Phi_c u_e) \cdot \Psi_e^{-1}$. Φ_b is the base frequency and Φ_c is the scaling factor of how much sampling frequency needs to be increased per each flow unit.

Values of sampling frequency determined by the described method are presented in Figure 6 as a histogram.



Figure 6. Histogram of sampling frequencies in the network.

4.4. Experiments

This section presents results of numerical experiments obtained with MIP solver and constant parameters presented in Table 4. Our experiments were divided into two cases:

- Case A—simplified dispersion model data—as explained in Section 4.2—with sampling depending on flow and pipe size;
- Case B—dispersion model data based on simulated discharges—as explained in Section 4.1—with sampling depending on flow and pipe size.

Table 4. Values of the used parameters.

Parameter	Value
A	7
В	3
Γ_e	5
Ω_e	10^{6}
Φ_h	1/60
Φ_c	1/60
Θ	10 ⁶

Each case was tested with $\Pi = 0.1, 0.2, \ldots, 0.9, 1.0$ to determine how the cost changes when the constraint on how many pollution sources have to be covered is changed. The obtained results are presented in Table 5 and in Figure 7 for dataset 1 and in Table 6 and Figure 8 for dataset 2.

	Cost [Cost Units]								
Coverage [%]	EC 2000 3:00	EC 3000 3:00	EC 4000 3:00	EC 2000 8:00	EC 3000 8:00	EC 4000 8:00			
10	15	15	15	15	15	15			
20	15	15	15	15	15	30			
30	15	30	30	15	30	45			
40	30	30	30	30	45	60			
50	30	45	45	30	60	90			
60	30	60	60	45	75	120			
70	45	75	90	60	120	165			
80	60	90	135	90	183	255			
90	75	135	213	120	303	393			
100	168	303	471	250	600	750			

Table 5. Cost function values for the test scenarios of dataset 1.



Figure 7. Optimal cost of IoT equipment deployment for dataset 1. Sampling frequency in a given pipe depends on the flow and the size of the pipe.

		Cost [Co	st Units]	
Coverage [/6]	k = 10	k = 20	k = 30	k = 40
10	36	24	21	21
20	87	48	36	33
30	144	69	54	48
40	210	102	69	69
50	285	132	93	87
60	369	171	126	111
70	480	222	168	144
80	645	285	213	186
90	915	378	285	261
100	1563	744	597	597

 Table 6. Cost function values for the test scenarios of dataset 2.



Figure 8. Optimal cost of IoT equipment deployment for dataset 2. Sampling frequency in a given pipe depends on the flow and the size of the pipe.

Both Figures 7 and 8 show an exponential increase of the cost for an increase in the demanded percentage of the sub-catchment area coverage. For instance, for the scenario when the threshold is set to $Q_3 = 4 \text{ mS/cm}$ and there are low flow conditions (f_L), a reduction of the cost of 47.6%—i.e., from 750 cost units to 393 cost units—can be achieved when relaxing the covered area from 100% to 90%. Similar relative cost reductions can be achieved at 90% coverage for all other five evaluated scenarios in each case.

Such results demonstrate that a wide area coverage is economically feasible for endusers—Law Enforcement Agencies and Environmental Agencies (LEAEA)—interested in monitoring an urban area, if the requirement of covering the whole sub-catchment area is relaxed. From these results, we conjecture that end-users may attempt to select for omission in the planning 10% of sources with a low probability of illegal discharges with the aim of reducing the cost of deployment by almost one half. This conjecture shall be studied in further work.

Figures 9 and 10 show the computational efficiency of the proposed method. Figure 9 shows the time as a function of the percentage coverage of the network for a representative case of the experiment shown in Figure 8. It should be emphasised that the computational time is satisfactory, with the cases between 40% and 80% coverage taking the most computational time.

On the other hand, Figure 10 shows convergence curve as a function of gap and the number of iterations. The gap reflects the difference between the best known bound and the objective value of the best solution produced by a particular algorithm.

Some statistical results concerning space utilization in the edges for both data-set scenarios are also presented in Appendix A.



Figure 9. Computation time. Sampling frequency in a given pipe depends on flow and the size of the pipe.



Figure 10. Convergence of the MIP method.

5. Conclusions

This work has addressed the problem of coverage in the sewage network. A model is proposed that provides a coverage problem in a sewer network and at the same time optimises network infrastructure resources such as Micromole rings with modules including sensors and batteries. We proposed the mixed integer programming method, which guarantees to find an optimal solution. In the experiments we used an example of a wideranging realistic sewage network from a big-sized city. The method we proposed proved to be effective, giving optimal results in a reasonable computational time.

The convergence curves show an exponential increase in cost for an increase in the desired percentage of coverage of the sub-catchment area. These results show that a wide range of coverage is economically feasible for end users. Based on these results, we conjecture that end-users may try to select up to a dozen percent of sources with low probability of illicit discharges for omission in planning in order to reduce the cost of deployment by almost half. This idea will be the subject of our further research in this area. We plan to develop a model and cost function to locate a potential source of pollutant discharge in the sewer network. We also plan to use evolutionary and bee algorithms if the computation time is long.

Author Contributions: Conceptualization, S.K., F.S.D. and A.S.; methodology, S.K.; software, A.S.; validation, S.K., F.S.D. and A.S.; formal analysis, S.K.; data curation, F.S.D.; writing-original draft preparation, S.K.; writing-review and editing, F.S.D., S.K. and A.S.; visualization, A.S.; supervision, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the H2020 SYSTEM project, which had received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 787128.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thanks the contribution of Steffen Krausse and Omar Shehata from Bundeswehr University Munich, Germany, for the discharge simulation performed by the ++SYSTEM Isar in Section 4.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

WSN	Wireless Sensor Networks
IoT	Internet of Things
MIP	Mixed Integer Programming
WWTP	WasteWater Treatment Plants
EC	Electrical Conductivity
NSGA	Non-Dominated Sorting Genetic Algorithm
SBN	Simulated Binary Crossover
AI	Artificial Intelligence
DFS	Depth-First Search
gap	Difference between current best integer solution and optimal value of LP relaxation
LEAEA	Law Enforcement Agencies and Environmental Agencies
CPLEX	Mixed Integer Programming solver.

Appendix A

In the appendix, aggregated statistics of cross-sectional area utilization of pipes by sensors and batteries, or simply edge space utilization, per test scenario are included. Only edges with $\gamma_e = 1$ were considered in the statistics. In all cases $\alpha_e = 1$, so statistics of α_e were omitted in the tables. Edge utilization is measured as the ratio between the number of slots used by batteries and sensors and the total number of slots available in the given edge. Edge utilization means the number of edges with $\gamma_e = 1$.

For dataset 1 it can be concluded that for cases with hour 8:00, edge utilization is greater than for cases with hour 3:00. Space utilization is lower for 8:00, however. The statistics are presented in Table A1. For dataset 2 it can be concluded that the greater the k value is, the lower the edge utilization is. The same observation can be made for average slot (space) utilization—the greater the *k* value, the lower the space utilization. In addition, the greater the coverage percentage, the greater the space utilization is. The statistics are presented in Table A2. For both datasets it can be observed that the greater the coverage percentage is, the greater the edge utilization is.

	Percentage Coverage	Utilized	Utilized β_c			Space Utilization [%]		
Scenario	[%]	Edges	Min	Max	Mean	Min	Max	Mean
	10	1	1	1	1.00	13.33	13.33	13.33
	20	1	1	1	1.00	13.33	13.33	13.33
	30	1	1	1	1.00	13.33	13.33	13.33
	40	2	1	1	1.00	13.33	13.33	13.33
	50	2	1	1	1.00	13.33	13.33	13.33
EC2000 03:00	60	2	1	1	1.00	13.33	13.33	13.33
	70	3	1	1	1.00	13.33	13.33	13.33
	80	4	1	1	1.00	13.33	13.33	13.33
	90	5	1	1	1.00	13.33	13.33	13.33
	100	11	1	2	1.09	13.33	20.00	13.94
	10	1	1	1	1.00	13.33	13.33	13.33
	20	1	1	1	1.00	13.33	13.33	13.33
	30	1	1	1	1.00	13.33	13.33	13.33
	40	2	1	1	1.00	13.33	13.33	13.33
EC2000 08.00	50	2	1	1	1.00	13.33	13.33	13.33
EC2000 08:00	60	3	1	1	1.00	13.33	13.33	13.33
	70	4	1	1	1.00	13.33	13.33	13.33
	80	6	1	1	1.00	13.33	13.33	13.33
	90	8	1	1	1.00	13.33	13.33	13.33
	100	9	1	1	1.00	13.33	13.33	13.33
	10	1	1	1	1.00	13.33	13.33	13.33
	20	1	1	1	1.00	13.33	13.33	13.33
	30	2	1	1	1.00	13.33	13.33	13.33
	40	2	1	1	1.00	13.33	13.33	13.33
EC2000 02.00	50	3	1	1	1.00	13.33	13.33	13.33
EC3000 03:00	60	4	1	1	1.00	13.33	13.33	13.33
	70	5	1	1	1.00	13.33	13.33	13.33
	80	6	1	1	1.00	13.33	13.33	13.33
	90	9	1	1	1.00	13.33	13.33	13.33
	100	20	1	2	1.05	13.33	40.00	15.00
	10	1	1	1	1.00	13.33	13.33	13.33
	20	1	1	1	1.00	13.33	13.33	13.33
	30	2	1	1	1.00	13.33	13.33	13.33
	40	3	1	1	1.00	13.33	13.33	13.33
EC2000 08:00	50	4	1	1	1.00	13.33	13.33	13.33
EC3000 08.00	60	5	1	1	1.00	13.33	13.33	13.33
	70	8	1	1	1.00	13.33	13.33	13.33
	80	12	1	2	1.08	13.33	20.00	13.89
	90	20	1	2	1.05	13.33	20.00	13.67
	100	23	1	2	1.05	13.33	20.00	13.67
	10	1	1	1	1.00	13.33	13.33	13.33
	20	1	1	1	1.00	13.33	13.33	13.33
	30	2	1	1	1.00	13.33	13.33	13.33
	40	2	1	1	1.00	13.33	13.33	13.33
FC4000 03:00	50	3	1	1	1.00	13.33	13.33	13.33
10100 00.00	60	4	1	1	1.00	13.33	13.33	13.33
	70	6	1	1	1.00	13.33	13.33	13.33
	80	9	1	1	1.00	13.33	13.33	13.33
	90	14	1	2	1.07	13.33	20.00	13.81
	100	31	1	2	1.06	13.33	40.00	15.48

 Table A1. Aggregated statistics of space utilization in the edges of dataset 1 test scenarios.

	Percentage Coverage	Utilized		βe		Space Utilization [%]		
Scenario	[%]	Edges	Min	Max	Mean	Min	Max	Mean
	10	1	1	1	1.00	13.33	13.33	13.33
	20	2	1	1	1.00	13.33	13.33	13.33
	30	3	1	1	1.00	13.33	13.33	13.33
	40	4	1	1	1.00	13.33	13.33	13.33
EC4000.08.00	50	6	1	1	1.00	13.33	13.33	13.33
EC4000 06:00	60	8	1	1	1.00	13.33	13.33	13.33
	70	11	1	1	1.00	13.33	13.33	13.33
	80	17	1	1	1.00	13.33	13.33	13.33
	90	26	1	2	1.04	13.33	20.00	13.59
	100	31	1	2	1.04	13.33	20.00	13.59

Table A1. Cont.

 Table A2. Aggregated statistics of space utilization in the edges of dataset 2 test scenarios.

C	Percentage	Utilized		β_e		Space	on [%]	
Scenario	Coverage [%]	Edges	Min	Max	Mean	Min	Max	Mean
	10	2	1	2	2.00	20.00	20.00	20.00
	20	5	1	3	1.80	13.33	26.67	18.67
	30	7	1	7	2.86	13.33	53.33	25.71
	40	11	1	7	2.36	13.33	53.33	22.42
k = 10	50	15	1	7	2.33	13.33	53.33	22.22
$\kappa = 10$	60	20	1	7	2.15	13.33	53.33	21.00
	70	26	1	7	2.15	13.33	85.71	22.78
	80	35	1	7	2.14	13.33	100.00	24.16
	90	48	1	7	2.35	13.33	100.00	30.26
	100	87	1	7	1.99	13.33	100.00	31.26
	10	1	4	4	4.00	33.33	33.33	33.33
	20	2	1	7	4.00	13.33	53.33	33.33
	30	3	1	7	3.67	13.33	53.33	31.11
	40	5	1	4	2.80	13.33	33.33	25.33
k - 20	50	6	1	7	3.33	13.33	53.33	28.89
$\kappa = 20$	60	8	1	7	3.12	13.33	53.33	27.50
	70	11	1	7	2.73	13.33	53.33	24.85
	80	14	1	7	2.79	13.33	53.33	25.24
	90	20	1	7	2.30	13.33	53.33	22.00
	100	39	1	7	2.36	13.33	85.71	29.80
	10	1	3	3	3.00	26.67	26.67	26.67
	20	2	1	3	2.00	13.33	26.67	20.00
	30	2	3	7	5.00	26.67	53.33	40.00
	40	3	1	7	3.67	13.33	53.33	31.11
k - 30	50	4	1	7	3.75	13.33	53.33	31.67
$\kappa = 50$	60	6	1	7	3.00	13.33	53.33	26.67
	70	8	1	7	3.00	13.33	53.33	26.67
	80	10	1	7	3.10	13.33	53.33	27.33
	90	15	1	7	2.33	13.33	53.33	22.22
	100	33	1	7	2.03	13.33	66.67	27.58
	10	1	3	3	3.00	26.67	26.67	26.67
	20	1	7	7	7.00	53.33	53.33	53.33
	30	2	1	7	4.00	13.33	53.33	33.33
	40	3	1	7	3.67	13.33	53.33	31.11
k = 40	50	4	1	7	3.25	13.33	53.33	28.33
$\kappa = 40$	60	5	1	7	3.40	13.33	53.33	29.33
	70	7	1	7	2.86	13.33	53.33	25.71
	80	9	1	7	2.89	13.33	53.33	25.93
	90	14	1	7	2.21	13.33	53.33	21.43
	100	33	1	7	2.03	13.33	66.67	27.58

References

- Hauser, F.M.; Hulshof, J.W.; Rößler, T.; Zimmermann, R.; Pütz, M. 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 detection. *Drug Test. Anal.* 2018, 10, 1368–1382. [PubMed]
- Cormier, G.J. Alkaline Cleaning. In Surface Engineering; Cotell, C.M., Sprague, J.A., Smidt, F.A., Eds.; ASM International: Materials Park, OH, USA, 1994; pp. 18–20. [CrossRef]
- Ghasemipanah, K. Treatment of ion-exchange resins regeneration wastewater using reverse osmosis method for reuse. *Desalin. Water Treat.* 2013, *51*, 5179–5183. [CrossRef]
- Consortium, H.M. Micromole—Sewage Monitoring System for Tracking Synthetic Drug Laboratories. Available online: http://www.micromole.eu (accessed on 17 October 2019).
- Consortium, H.S. H2020 SYSTEM—Synergy of Integrated Sensors and Technologies for Urban Secured Environment. Fact Sheet Available at EC Website. Under Project ID 787128. Available online: https://cordis.europa.eu/project/rcn/220304/factsheet/en (accessed on 6 September 2022).
- De Vito, S.; Fattoruso, G.; Esposito, E.; Salvato, M.; Agresta, A.; Panico, M.; Leopardi, A.; Formisano, F.; Buonanno, A.; Delli Veneri, P.; et al. A Distributed Sensor Network for Waste Water Management Plant Protection. In *Convegno Nazionale Sensori*; Andò, B., Baldini, F., Di Natale, C., Marrazza, G., Siciliano, P., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 303–314.
- Lepot, M.; Makris, K.F.; Clemens, F.H. Detection and quantification of lateral, illicit connections and infiltration in sewers with Infra-Red camera: Conclusions after a wide experimental plan. *Water Res.* 2017, 122, 678–691. [CrossRef] [PubMed]
- 8. Tan, F.H.S.; Park, J.R.; Jung, K.; Lee, J.S.; Kang, D.K. Cascade of One Class Classifiers for Water Level Anomaly Detection. *Electronics* 2020, *9*, 1012. [CrossRef]
- Tashman, Z.; Gorder, C.; Parthasarathy, S.; Nasr-Azadani, M.M.; Webre, R. Anomaly Detection System for Water Networks in Northern Ethiopia Using Bayesian Inference. *Sustainability* 2020, 12, 2897. [CrossRef]
- Zhang, D.; Heery, B.; O'Neil, M.; Little, S.; O'Connor, N.E.; Regan, F. A Low-Cost Smart Sensor Network for Catchment Monitoring. *Sensors* 2019, 19, 2278. [CrossRef] [PubMed]
- 11. Perfido, D.; Messervey, T.; Zanotti, C.; Raciti, M.; Costa, A. Automated Leak Detection System for the Improvement of Water Network Management. *Proceedings* **2016**, *1*, 28. [CrossRef]
- 12. Rojek, I.; Studzinski, J. Detection and Localization of Water Leaks in Water Nets Supported by an ICT System with Artificial Intelligence Methods as a Way Forward for Smart Cities. *Sustainability* **2019**, *11*, 518. [CrossRef]
- Ji, H.; Yoo, S.; Lee, B.J.; Koo, D.; Kang, J.H. Measurement of Wastewater Discharge in Sewer Pipes Using Image Analysis. *Water* 2020, 12, 1771. [CrossRef]
- 14. Kuchmenko, T.A.; Lvova, L.B. A Perspective on Recent Advances in Piezoelectric Chemical Sensors for Environmental Monitoring and Foodstuffs Analysis. *Chemosensors* **2019**, *7*, 39. [CrossRef]
- Pisa, I.; Santín, I.; Vicario, J.; Morell, A.; Vilanova, R. ANN-Based Soft Sensor to Predict Effluent Violations in Wastewater Treatment Plants. *Sensors* 2019, 19, 1280. [CrossRef]
- 16. Drenoyanis, A.; Raad, R.; Wady, I.; Krogh, C. Implementation of an IoT Based Radar Sensor Network for Wastewater Management. Sensors 2019, 19, 254. [CrossRef]
- 17. Ma, J.; Meng, F.; Zhou, Y.; Wang, Y.; Shi, P. Distributed Water Pollution Source Localization with Mobile UV-Visible Spectrometer Probes in Wireless Sensor Networks. *Sensors* **2018**, *18*, 606. [CrossRef]
- Desmet, C.; Degiuli, A.; Ferrari, C.; Romolo, F.; Blum, L.; Marquette, C. Electrochemical Sensor for Explosives Precursors' Detection in Water. *Challenges* 2017, 8, 10. [CrossRef]
- 19. Solano, F.; Krause, S.; Wöllgens, C. An Internet-of-Things Enabled Smart System for Wastewater Monitoring. *IEEE Access* 2022, 10, 4666–4685. [CrossRef]
- 20. Consortium, S. SIMONA—Sistema Integrato di Competenze per il MONitoraggio, la Protezione e il Controllo Delle Infrastrutture Idriche, Fognarie e Ambientali. Available online: http://www.progettosimona.it (accessed on 15 March 2021).
- Banik, B.; Alfonso, L.; Torres, A.; Mynett, A.; Di Cristo, C.; Leopardi, A. Optimal Placement of Water Quality Monitoring Stations in Sewer Systems: An Information Theory Approach. *Procedia Eng.* 2015, 119, 1308–1317. [CrossRef]
- 22. Banik, B.; Alfonso, L.; Di Cristo, C.; Leopardi, A. Greedy Algorithms for Sensor Location in Sewer Systems. *Water* 2017, *9*, 856. [CrossRef]
- 23. Banik, B.K.; Alfonso, L.; Cristo, C.D.; Leopardi, A.; Mynett, A. Evaluation of Different Formulations to Optimally Locate Sensors in Sewer Systems. *J. Water Resour. Plan. Manag.* **2017**, *143*, 04017026. [CrossRef]
- 24. Ali, H.; Osmani, S.; Banik, B. Integrating fuzzy logic with Pearson correlation to optimize contaminant detection in water distribution system with uncertainty analyses. *Environ. Monit. Assess.* **2019**, *191*, 441. [CrossRef]
- 25. Alfonso, L.; He, L.; Lobbrecht, A.; Price, R. Information theory applied to evaluate the discharge monitoring network of the Magdalena River. *J. Hydroinformatics* **2013**, *15*, 211–228. [CrossRef]
- 26. Alfonso, L.; Lobbrecht, A.; Price, R. Information theory–based approach for location of monitoring water level gauges in polders. *Water Resour. Res.* **2010**, *46*, 3. [CrossRef]

- 27. Alfonso, L.; Lobbrecht, A.; Price, R. Optimization of water level monitoring network in polder systems using information theory. *Water Resour. Res.* 2010, *46*, 12. [CrossRef]
- Sikorski, A.; Kozdrowski, S.; Donado, F.S. IoT Device Deployment for Optimal Wastewater Network Coverage. In Proceedings of the 2021 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Hvar, Croatia, 23–25 September 2021; pp. 1–6. [CrossRef]
- Tossa, F.; Abdou, W.; Ezin, E.C.; Gouton, P. Improving Coverage Area in Sensor Deployment Using Genetic Algorithm. In *Computational Science—ICCS 2020*; Krzhizhanovskaya, V.V., Závodszky, G., Lees, M.H., Dongarra, J.J., Sloot, P.M.A., Brissos, S., Teixeira, J., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 398–408.
- Hanh, N.T.; Binh, H.T.T.; Hoai, N.X.; Palaniswami, M.S. An efficient genetic algorithm for maximizing area coverage in wireless sensor networks. *Inf. Sci.* 2019, 488, 58–75. [CrossRef]
- Hanh, N.T.; Nam, N.H.; Binh, H.T.T. Particle Swarm Optimization Algorithms for Maximizing Area Coverage in Wireless Sensor Networks. In *Proceedings of the SAI Intelligent Systems Conference (IntelliSys)* 2016; Bi, Y., Kapoor, S., Bhatia, R., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 893–904.
- 32. Buras, M.P.; Solano Donado, F. Identifying and Estimating the Location of Sources of Industrial Pollution in the Sewage Network. *Sensors* **2021**, 21, 3426. [CrossRef]
- 33. Chachuła, K.; Nowak, R.; Solano, F. Pollution Source Localization in Wastewater Networks. Sensors 2021, 21, 826. [CrossRef]
- 34. Chachuła, K.; Słojewski, T.M.; Nowak, R. Multisensor Data Fusion for Localization of Pollution Sources in Wastewater Networks. Sensors 2022, 22, 387. [CrossRef]
- Fourer, R.; Gay, D.M.; Kernighan, B.W. AMPL: A Mathematical Programing Language. In Algorithms and Model Formulations in Mathematical Programming; Wallace, S.W., Ed.; Springer: Berlin/Heidelberg, Germany, 1989; pp. 150–151.
- Kozdrowski, S.; Żotkiewicz, M.; Wnuk, K.; Sikorski, A.; Sujecki, S. A Comparative Evaluation of Nature Inspired Algorithms for Telecommunication Network Design. *Appl. Sci.* 2020, 10, 6840. [CrossRef]
- 37. Tandler.com. Software for Water Management ++ Systems Isar. Available online: www.tandler.com (accessed on 21 November 2021).