Journal of Information Assurance and Security 4 (2009) 500-508

Security Issues in Drinking Water Distribution Networks

Demetrios G. Eliades[1], Marios M. Polycarpou[1] and Bambos Charalambous[2]

[1]KIOS Research Center for Intelligent Systems and Networks,
Dept. of Electrical and Computer Engineering, University of Cyprus
P.O. Box 20537, 1678 Nicosia, Cyprus
{eldemet,mpolycar}@ucy.ac.cy

[2]Water Board of Lemesos
66 Franklin Roosevelt Avenue, Lemesos, 3602, Cyprus
bambos@wbl.com.cy

Abstract: Critical infrastructure security will be a key challenge in the years ahead. From a system-theoretic viewpoint, there is a need to develop rigorous design and analysis tools, suitable for addressing the security problems in the various critical infrastructures. In this work, we develop a mathematical framework suitable for addressing security issues in drinking water distribution networks. In addition, we investigate the problem of evaluating the most vulnerable node locations in the network in order to physically secure them, as well as that of determining a suitable set of locations and/or sampling locations and times, so that the effects of a contamination fault can be minimized. This work contributes to the research by presenting a mathematical formulation of the problem suitable for solving the sensor placement and manual sampling simultaneously, under realistic conditions. The optimization problem is solved using various optimization and evolutionary computation techniques. To illustrate the solution methodology, we present results based on a real water distribution network.

Keywords: Water distribution networks, Water systems, Security, Fault diagnosis framework, Sensor placement

1. Introduction

Critical infrastructure security has received significant attention within the past few years, especially due to various terrorist acts around the globe. Critical infrastructures are defined by the European Commission[1] as those systems and assets which are essential for the operation of vital societal functions, and whose disruption of operation would have a significant negative impact on society. They are comprised of both physical and information technology facilities, networks and services. These may include energy and telecommunication networks, financial services, transport systems, health delivery services as well as safe food and drinking water supply. Addressing security issues for the critical infrastructures is a complex task; for example, a large number of different policy makers are involved for each infrastructure, and some infrastructures may be shared by many countries with different regulations. In addition, various parameters can be considered for measuring the damage caused on a critical infrastructure due to a fault, such as economic losses or the number of people affected.

Protecting the critical infrastructure from malicious acts, natural hazards and accidental faults requires the development of fault diagnosis and security frameworks; this involves the synergy of various research fields, such as modelling, control, risk management and optimization.

In the present work, we examine a number of security issues related with a certain type of distribution networks which are part of the critical infrastructure, and in specific with the drinking water distribution systems. Distribution networks can be represented in an abstract manner as a directed graph, with nodes at certain physical or virtual locations which are connected to some other nodes by a line, or a transportation path. For example, a power distribution network is comprised of buses and cables, an underground metro network is comprised of stations and tube lines and a water distribution network is comprised of pipes and junctions.

Consider a simple distribution network as in Fig. 1. When a fault occurs at a certain node, the fault may propagate to the other nodes with some time delay, following the direction of the flow. For example, a fault in node v_2 would propagate, affecting nodes $\mathcal{F}_{v_s} = \{v_3, v_4, v_5, v_6, v_7\}$, if the fault is not detected and accommodated in time.

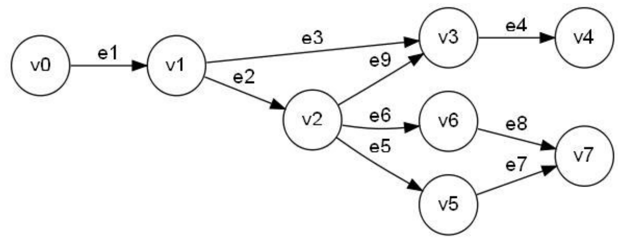

Figure 1. A simple transportation network.

In the case where the graph in Fig. 1 corresponds to a drinking water distribution network, each node and each directed edge would have a time-varying weight characterizing the volume of water consumed at that neighbourhood, and the transport delays, respectively. A quality fault in the

[1] COM(2008) 676, Proposal for a Council Decision on a Critical Infrastructure Warning Information Network (CIWIN).

Received June 10, 2009
1554-1010 $ 03.50 Dynamic Publishers, Inc.

system, i.e. a contaminant substance which has been injected somehow at a certain node, will propagate along the water flow and may be ingested by the consumers served in its path. Health regulations require water utilities to monitor the quality of water by collecting data (either by manual sampling or by using on-line sensors) at various locations in the network; these would serve to determine the chemical concentrations of contaminants, disinfectants (such as chlorine), or other chemical variables.

The standard approach is for the water utility to perform manual water sampling for quality analysis, at various nodes within a network, and for a few times during the day [1]. Advanced quality sensors installed permanently at fixed locations in the network, are currently being used by some water utilities for automated sampling; however, the high cost required to purchase, install and maintain quality sensors prohibits water utilities from installing them everywhere. In practice, a water utility would install a small number of water quality sensors in the distribution network, at certain "strategic" locations.

In the present work we present a mathematical framework suitable for addressing several security problems related with drinking water distribution networks. In specific, we examine the problem of determining which parts of the distribution network have a high-risk for contamination faults, as well as the problem of computing where to install water quality sensors or/and where and when to conduct manual sampling, such that the damage due to a contamination fault is minimized according to certain metrics. The proposed approach is illustrated on a real water distribution network.

In Section 2, we provide a description of possible hydraulic and quality faults in water distribution networks, and present background information on research related to the sensor placement problem. In Section 3, we present the mathematical framework suitable for addressing security issues in water distribution networks. In Section 4 we discuss various security issues and propose some solution methodologies for minimizing the quality fault risk. In Section 5 we demonstrate the methodology with an illustrative example. Finally, in Section 6, we present some final conclusions and future work.

2. Description of Water Systems

2.1 Faults in Water Systems

There are two categories of faults affecting water distribution systems: hydraulic faults and quality faults; the former are sometimes a consequence of the latter. Following, is a brief description of hydraulic and quality faults in water systems.

2.1.1 Hydraulic Faults

Hydraulic faults may be due to leakages, pipe bursts, blocked pipes or malfunctioning pumps and valves [1]. A burst is an abrupt fault requiring immediate attention, and is usually easy to isolate. On the other hand, a leakage is an incipient fault, which is, in general, difficult to detect and isolate. Hydraulic faults can be detected by analyzing flow and pressure data acquired from SCADA [1].

Water loss due to hydraulic faults imposes a severe economic burden on the water utilities, while water supplies are reduced. In addition, hydraulic faults may cause quality faults, since contaminants may infiltrate the distribution system.

2.1.2 Quality Faults

Quality faults may occur due to the contamination of water by certain substances, usually chemical, biological or radioactive, which travel along the flow of water and may exhibit decay or growth dynamics. Certain disinfectants, such as chlorine, are used in prescribed concentrations, in order to maintain the quality of drinking water by preventing bacteria growth and neutralizing chemical agents [1]. However, these disinfectants in large concentrations have a similar effect as contaminants, since they can be a threat to human health.

In addition to natural and accidental contamination of a water distribution system, contaminants may be injected intentionally in the network. These types of quality faults are related to the security issue of water resources, which is becoming more and more important. In the worst case scenario, the contamination will be part of a well informed attack, seeking to cause economic losses and affect the served population's health dramatically. Therefore, although an intentional contamination can occur practically anywhere in the network, it may well occur in a part of the network, from which a large part of the population is exposed. Apart from the location, the time of fault occurrence also plays a significant role in the damage caused. This is due to the varying contaminant's transport time delays, caused by the propagation dynamics. Indeed, contaminant propagation modelling and security issues have received significant attention during the last decade, especially after the rise of terrorist attacks around the globe.

2.2 Sensor Placement for Water Quality Monitoring

The problem of where to install sensors in order to keep certain objectives and constraints satisfied within the network, has been examined in various research disciplines such as operational research, systems theory and control, combinatorial optimization, etc.

The "Set Covering" has been one of the first mathematical formulations of the problem, having been applied in various fields [2]. For the problem of where to install some facilities, a topological graph is considered and a subset of nodes is computed using an integer optimization program, so that each node is adjacent to at least one solution node. A related approach is the "Maximal Covering" formulation described in [3] for computing a set of nodes which maximize the population served in an area within a certain distance. A similar formulation to the Maximal Covering was considered in [4] for selecting the locations where to install quality sensors in water distribution systems, in order to inspect the maximum volume of water consumed. The authors proposed a scenario-based approach which segments a day into time-periods, corresponding to different flow patterns, and the optimization was solved for all scenarios simultaneously. Their approach did not address any security issues.

A mathematical formulation suitable for the security issues related to the location selection is the "p-median" [5]. A similar formulation was examined in [6], [7] for the water distribution systems. By considering a number of contamination scenarios and their impacts, the authors formulated a mathematical program to minimize the average "contamination impact". A multi-objective extension was further examined in [8]; for solving the mixed-integer optimization program, however, significant computational power was required.

Within the water resources management community, the design competition of the "Battle of the Water Sensor Networks (BWSN)" in 2006, instigated significant research interest and discussion on security issues of water distribution systems [9]; the task was to find sets of locations where to install sensors using two real benchmark networks, so that a number of objectives are optimized under various fault scenarios. The majority of the participants formulated a multi-objective integer optimization program, such as [10].

3. Mathematical Framework

In this section we present a mathematical framework, suitable for analyzing certain security issues in water distribution networks.

3.1 Advection and Impact Dynamics

Consider a water distribution network comprised of pipes, junctions and water storages. The topology of this network can be represented as a graph with edges corresponding to pipes, and m nodes corresponding to junctions and water storages; we assume that an injection of a substance can occur at any one of these m nodes.

For modelling purposes, each pipe in the network is *a priori* virtually segmented into a number of finite volume cells. Let n be the total number of the nodes and the finite volume cells considered in the network. Let $x_i(k)$ denote the concentration of a certain contaminant at discrete time k within each node and finite volume cell, for $i = 1, ..., m, ..., n$. The vector $x(k) = [x_1(k), ..., x_m(k), ..., x_n(k)]^\top$ is the state of the contaminant concentration dynamics. The set of all node indices is $\mathcal{V} = \{1, 2, ..., m\}$.

The advection-reaction equations [11] describing the propagation of a contaminant in a water distribution network can be expressed in a state-space formulation:

$$\begin{aligned} x(k+1) &= A(k)x(k) + \zeta(x(k)) + F\varphi(k) \\ y(k) &= Cx(k), \end{aligned}$$

where $A(k)$ is an $n \times n$ matrix which characterizes the advection dynamics, and $\zeta : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a function which describes the reaction dynamics of the contaminant. For m possible fault locations (i.e. at the nodes), let F be an $n \times m$ matrix describing the locations of a fault; function $\varphi : \mathbb{N}^+ \mapsto \mathbb{R}^m$ describes the change in the contaminant concentration at the contamination source. We can assume that the fault function $\varphi(k)$ can be represented through z linearly parameterized basis functions $p(k) = [p_1(k), ..., p_z(k)]^\top$, such that $\varphi(k) = \Theta p(k)$, where $\Theta \in \mathbb{R}^{m \times z}$. The (i, j)-th element of Θ, denoted as $\Theta_{(i,j)}$ represents the amplitude of the basis function $p_j(k)$ which is added to the state $x_i(k)$.

The output vector $y(k)$ corresponds to state measurements, which are monitored using quality sensors. In specific, for M_s quality sensors, C is a binary matrix $C \in \{0, 1\}^{M_s \times n}$, such that its (i, j)-th element is $C_{(i,j)} = 1$ when the i-th quality sensor measures the j-th state, and zero when there is no quality sensor, $C_{(i,j)} = 0$. This matrix can be chosen following a security-oriented methodology, as we will demonstrate in the next sections.

When a contamination fault occurs, the contaminant is propagated in the network and it may eventually reach the customers who will consume the outflow water. Let $\mathcal{W} \subseteq \mathcal{V}$ be the set of $N_w = |\mathcal{W}|$ node indices corresponding to locations which outflow water to consumers based on demand requests. For each "demand node" $w_i \in \mathcal{W}$, $i = 1, ..., N_w$, an impact value can be computed at each time step. The impact due to a contamination fault can be expressed by using epidemiological terms, e.g. how many people are infected, or by using economic terms, such as the losses cost. Other impact measures which can be considered include the consumed volume of contaminated water which exceeds a certain concentration threshold [9].

In general, the impact of a fault depends on the volume and concentration of the contaminated water consumed. This can be described by a dynamic equation, which we refer to as the *impact dynamics*. Let $\xi \in \mathbb{R}^{N_w}$ be the impact state vector which describes the "damage" caused at each demand node, at discrete time k, after a contaminant has been injected somewhere in the network. For $w_i \in \mathcal{W}$, and $i = 1, ..., N_w$, a state-space representation of the impact dynamics is given by

$$\begin{aligned} \xi_i(k+1) &= \xi_i(k) + f_\xi(x_{w_i}(k), d_{w_i}(k)) \\ \psi(k) &= f_\psi(\xi(k)), \end{aligned}$$

where $d_{w_i}(k)$ is the outflow demand (in $\frac{m^3}{s}$) at demand node w_i, and $f_\xi : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}^+$ is a non-negative function which characterizes the impact increase at each time step. The output impact vector $\psi(k)$ characterizes the overall impact, or "damage", caused by a certain contamination fault; function $f_\psi : \mathbb{R}^{N_w} \mapsto \mathbb{R}^+$ computes the overall impact.

In the case when N_ξ impact metrics are considered, then the impact dynamics for each metric are computed separately. Considering more than one impact metrics is useful for minimizing different types of "damages".

All the dynamics are in discrete time, with time step Δt. Hydraulic dynamics are approximately periodic; we define T_h as the discrete time length of one period (i.e. one day), given by $T_h = \frac{24\ hr}{\Delta t}$.

3.2 Contamination Scenarios

For some security-related problems in water distribution networks, it is useful to consider a number of representative fault scenarios, within a specific time-period, while the system is assumed to be operating under normal conditions. A scenario is comprised of two elements: a) the parameter-matrix Θ which describes the fault signal, and b) the time-delay t_d in shutting down the system.

Let $\mathcal{Z}$ be the set of all fault parameter matrices, such that

$$\mathcal{Z} = \left\{ \Theta \in \mathbb{R}^{m \times z} \mid \underline{\Theta}_{(i,j)} \leq \Theta_{(i,j)} \leq \overline{\Theta}_{(i,j)} \right\},$$

where $\underline{\Theta}_{(i,j)}, \overline{\Theta}_{(i,j)}$ are prespecified lower and upper bounds on the (i,j)-th element of Θ, for $i = 1, ..., m$ and $j = 1, ..., z$. In addition we define $\mathcal{T}_d$ as a set of various discrete time delays considered, for stopping the system after a fault has been detected. We define the set

$$\mathcal{H}^* = \{(\Theta, t_d) \mid \Theta \in \mathcal{Z}, t_d \in \mathcal{T}_d\}$$

as the set of all scenarios two-tuples. This set can be extremely large; a finite subset $\mathcal{H} \subset \mathcal{H}^*$ can be considered with $N_h = |\mathcal{H}|$ realistic fault scenarios, selected through grid or random sampling.

3.3 Quality Sampling Measurements

For the security problems examined in this work, we assume that from the set of all nodes indices $\mathcal{V}$, a subset $\mathcal{V}_s \subseteq \mathcal{V}$ corresponds to candidate locations for installing sensors or for conducting manual sampling (at certain times).

It is useful to define $\mathcal{T}_s$ as the set of discrete time instances when manual sampling can be performed, within one day (e.g. during working hours).

From these, we define the two-tuples set $\mathcal{Q}_m$, corresponding to candidate sampling node indices and sampling times, given by

$$\mathcal{Q}_m = \{(s, t_s) \mid s \in \mathcal{V}_s, t_s \in \mathcal{T}_s\}.$$

In addition, we define the set $\mathcal{Q}_s$, of candidate sensing node indices where sensors can be installed, such that

$$\mathcal{Q}_s = \{(s, \emptyset) \mid s \in \mathcal{V}_s\},$$

where $\emptyset$ is the empty set. Finally, we define $\mathcal{Q} = \{\mathcal{Q}_s, \mathcal{Q}_m\}$ as the set of all feasible sensing nodes and manual sampling times.

3.3 Detection and System Stop Time

For each fault scenario, in order to evaluate the damage caused, it is useful to compute the time of detection by either a sensor installed at a node, or by manual sampling performed at a node.

In practice, if a fault is not detected using sensor technologies or manual sampling, it will propagate for some discrete time T_{other}, after its first occurrence, until it is detected through other means, such as customer complaints, hospitalizations, etc. In addition, we define a discrete maximum time T_{max}, so that for any fault scenario, its impact will not increase after that time. For simplicity, we assume that $T_{max} = \lambda T_h$, where $\lambda \in \mathbb{N}^+$ and T_h is the discrete time duration of a one-day period. It is useful to define $\mathcal{T} = \{1, 2, ..., T_{max}\}$ as the set of time steps for which the system dynamics are simulated, as well as the period index set $\Lambda = \{0, 1, ..., \lambda\}$.

In order to compute the fault detection and system stop time, we define the function $g : \mathcal{H} \times \mathcal{Q} \times \mathbb{R}^+ \mapsto \mathcal{T}$. Therefore, for a scenario fault $h \in \mathcal{H}$, a sampling $q \in \mathcal{Q}$ and a concentration threshold $\epsilon \in \mathbb{R}^+$ above which a contaminant is assumed to be detected, we construct the function $g(h, q, \epsilon)$, which maps its parameters to a time step at which the fault has been detected and the system has been stopped.

In the following, the measurement of state i at time k when a fault with parameter matrix Θ occurs, will be denoted as $x_i(k; \Theta)$. For $h = (\Theta, t_d) \in \mathcal{H}$ and $q = (s, t_s) \in \mathcal{Q}$, the fault detection and system stop-time function is given by

$$g(h, q, \epsilon) = \begin{cases} \min\{\mathcal{T}_1\} + t_d & \text{if } \mathcal{T}_1 \notin \emptyset \text{ and } t_s = \emptyset \\ \min\{\mathcal{T}_2\} + t_d & \text{if } \mathcal{T}_2 \notin \emptyset \text{ and } t_s \neq \emptyset \\ \min\{\mathcal{T}_3\} + T_{other} + t_d & \text{if } \mathcal{T}_1 \in \emptyset \text{ or } \mathcal{T}_2 \in \emptyset \end{cases}$$

where

$$\begin{aligned} \mathcal{T}_1 &= \{k \mid x_s(k; \Theta) > \epsilon, k \in \mathcal{T}\} \\ \mathcal{T}_2 &= \{\eta T_h + t_s \mid x_s(\eta T_h + t_s; \Theta) > \epsilon, \eta \in \Lambda\} \\ \mathcal{T}_3 &= \{k \mid \|\Theta p(k)\| > 0, k \in \mathcal{T}\}. \end{aligned}$$

The fault detection and system halt time, in practice, depend on whether sensors or manual sampling are used at a certain node. The difference between them is that a water quality sensor will detect a fault on-line, as soon as a certain contaminant concentration threshold is surpassed; on the other hand, manual sampling will only detect a fault after a sample has been acquired at a certain time, and measured by trained personnel.

3.4 Overall Impact

The overall impact is the "damage" caused by a contamination fault, measured through certain impact metrics, such as the number of people infected or the volume of polluted water consumed. We compute the overall impact for each fault scenario with respect to each detection and system halt time. The impact of a fault with parameter matrix Θ at time k will be denoted as $\psi(k; \Theta)$.

For a certain impact metric, for $N_q = |\mathcal{Q}|$ and $N_h = |\mathcal{H}|$, we define Ω as the overall impact matrix, of size $N_h \times N_q$; the (i,j)-th element of this matrix is given by

$$\Omega_{(i,j)} = \psi(g(h, q, \epsilon); \Theta)$$

where $h = (\Theta, t_d) \in \mathcal{H}$, $q \in \mathcal{Q}$ and $\epsilon \in \mathbb{R}^+$.

As mentioned previously, a number of different impact metrics can be considered; for each, a different overall impact matrix is constructed according to our methodology. For N_ξ overall-impact metrics, we compute N_ξ overall-impact matrices, which belong to the set $\mathcal{J} = \{\Omega_1, ..., \Omega_{N_\xi}\}$.

4. Security Issues

4.1 Securing Neuralgic Locations

The first part of a security scheme in water distribution networks is to determine the locations in the water distribution network which could be considered as "high-risk" for contaminant injection, so that proper action is taken in order to secure them through physical means.

We first need to decide on a representative impact metric; from its corresponding overall-impact matrix $\Omega \in \mathcal{J}$, we will calculate the maximum scenario impact $\overline{\omega}$, given by

$$\overline{\omega} = \max_{i=1,...,N_h} \max_{j=1,...,N_q} \Omega_{(i,j)},$$

where $\Omega_{(i,j)}$ is the (i,j)-th element of Ω.

From this, we can compute the set of the worst-case scenarios; for example, the top 5% of set of fault indices is computed by

$$\mathcal{Y}_0 = \left\{ i \mid \max_{j=1,\dots,N_q} \Omega_{(i,j)} \geq 0.95\,\varpi,\ i \in \{1,2,\dots,N_h\} \right\}.$$

Thus, from the scenario indices included in $\mathcal{Y}_0$, it is easy to construct a list of the most neuralgic locations in the network which need to be physically secured. However, due to the nature of the water distribution networks, certain contamination faults can override physical security measures; it is thus imperative to use an extra layer of protection by installing a number of sensors at different locations in the network.

4.2 Minimum-Risk Quality Sensor Placement and Manual Sampling Scheduling

Deciding where to install quality sensors, and also where and when to perform manual sampling in a water distribution network, are problems with non-trivial solutions, which need to be addressed. By considering the mathematical formulation discussed in the previous section, the problem can be reduced to risk optimization.

4.2.1 Definition of Risk

In order to address the problem of security, it is important to have an understanding on what "risk" is and how it can be quantified. In general, risk is the possibility of an unpredictable future event that will result in losses, thus preventing the serving organization from meeting certain goals [12]. Risk has been examined in many fields and especially in the financial and operational research literature.

In financial practice, the most commonly used risk-objective is the "Value-at-Risk" (VaR), which represents the maximum loss with a certain confidence level over a time period. This metric, however, ignores the worst scenarios, which may be crucial in the case of intentional water contamination [12]. All in all, risk management provides useful tools and insights for the problem of security in critical infrastructure systems [8].

4.2.2 Solution Methodology

We assume that the task is to select M_s out of $N_{qs} = |Q_s|$ locations where sensors can be installed, and M_m out of $N_{qm} = |Q_m|$ two-tuples of locations and time instances, corresponding to where and when sampling can be performed. We define $\mathcal{L}_s = \{1,2,\dots,N_{qs}\}$ as the set of the candidate sensing node indices, and $\mathcal{L}_m = \{N_{qs}+1, N_{qs}+2,\dots,N_q\}$ as the set of candidate sampling location and time indices. In addition, we define $\mathbb{X}$ as the set of all solution combinations between these index sets, such that

$$\mathbb{X} = \left\{ \{X_s, X_m\} \mid X_s \in \mathcal{L}_s^{M_s}, X_m \in \mathcal{L}_m^{M_m} \right\}.$$

For N_f different objective functions, the optimization problem is constructed as

$$\mathcal{Y} = \operatorname*{argmin}_{\mathcal{X}\in\mathbb{X}} \{F_1(\mathcal{X};\Omega_a),\dots,F_{N_f}(\mathcal{X};\Omega_b)\},$$

where $F_i : \mathbb{X} \mapsto \mathbb{R}$ is an optimization function decided by the decision makers, and $\Omega_a, \Omega_b \in \mathcal{J}$. For $N_f = 1$, the solution $\mathcal{Y}$ corresponds to a single set of indices, whereas for $N_f > 1$, $\mathcal{Y}$ corresponds to one or more sets of indices.

For a certain $\mathcal{X} \in \mathbb{X}$ and a certain $\Omega \in \mathcal{J}$ the i-th objective function $F_i(\mathcal{X};\Omega)$ is computed through a risk function. In the next paragraphs we demonstrate some risk functions suitable for the security problem in water distribution networks.

In computing the risk objective functions, it is useful to define the scenario index set $\mathcal{G}^* = \{1,2,\dots,N_h\}$, corresponding to each of the N_h fault scenarios considered. In practice, a subset from the set of all scenarios could be neglected in the optimization process, depending on the risk-objective or on whether they are considered trivial with respect to their impact magnitude.

Average Impact: The average impact metric is suitable for optimizing reliability, in the case where contaminant injection can occur at any node with equal probability. This metric, however, has limitations when considering the security framework, since it does not take into sufficient consideration rare faults with extreme consequences. For a certain overall-impact matrix $\Omega \in \mathcal{J}$, a scenario index set $\mathcal{G} = \mathcal{G}^*$, and for a specific set of solution indices $\mathcal{X} \in \mathbb{X}$, the average impact across all faults is given by

$$F(\mathcal{X};\Omega) = \frac{1}{|\mathcal{G}|} \sum_{i\in\mathcal{G}} \min_{j\in\mathcal{X}} \Omega_{(i,j)}.$$

Maximum Impact: The maximum impact metric is used to reduce the effect of the most extreme fault, in terms of causing the most damage. This metric is useful from a security perspective; on the other hand, it does not take into consideration the fault frequency distribution, and in specific, the frequency of extreme faults. For a certain overall-impact matrix $\Omega \in \mathcal{J}$, a scenario index set $\mathcal{G} = \mathcal{G}^*$, and for a specific set of solution indices $\mathcal{X} \in \mathbb{X}$, the maximum impact across all faults is given by

$$F(\mathcal{X};\Omega) = \max_{i\in\mathcal{G}} \min_{j\in\mathcal{X}} \Omega_{(i,j)}.$$

4.2.3 Solving the Optimization Problem

For $N_f = 1$, the single-objective problem can be solved using an integer nonlinear optimization algorithm; a single solution is computed, i.e. $\mathcal{Y} \in \mathbb{X}$. In the multi-objective case, where $N_f > 1$, minimizing one objective may result in maximizing others; it is thus not possible to find one optimal solution which satisfies all objectives at the same time. It is possible, however, to find a subset of solutions $\mathcal{Y} \subset \mathbb{X}$, laying on a Pareto front, where each solution is no worse than the other. In the multi-objective case, a feasible solution $\mathcal{X}$ is called Pareto optimal if for a set of N_f objectives and for $\Omega_a, \Omega_b \in \mathcal{J}$, there exists no other feasible solution $\mathcal{X}'$ such that $F_i(\mathcal{X}', \Omega_a) \leq F_i(\mathcal{X};\Omega_a)$ with $F_j(\mathcal{X}';\Omega_b) < F_j(\mathcal{X};\Omega_b)$ for at least one j. Therefore, a solution is Pareto optimal if there is no other feasible solution which would reduce some objective function without simultaneously causing an increase in at least one other objective function [14, p.779].

Evolutionary computation approaches have proven useful for handling optimization problems with multiple objectives in the continuous and discrete domain. Some of these algorithms (see, for example, [15]) have already been applied to the sensor placement problem in various fields.

From a practical viewpoint, only one solution can be implemented. Let $Y \in \mathcal{Y}$ be the single solution selected through high-level reasoning by the decision makers; the set Y is comprised of M_s sensor location indices corresponding to the first M_s elements, and M_m manual sampling two-tuples corresponding to the remaining elements. In addition, the output matrix C which is part of the advection and reaction dynamics, is constructed using the set Y; in specific, for $i = 1, ..., M_s$, and for the node index j corresponding to the Y_i-th element in $\mathcal{V}_s$ (the set of location candidates), then $C_{(i,j)} = 1$, and zero otherwise.

5. Application Example

In this section, we present a simulation example to illustrate the formulation and the solution methodology for the security issues addressed in the present work, and in specific for sensor placement and manual quality sampling scheduling on a real drinking water distribution network, operating under realistic conditions.

5.1 Network Description

The network is comprised of 321 pipes connected to 198 junctions, 100 of which are used for water consumption. Disinfected water is supplied to the network from a single node. The structural characteristics are assumed to be known, i.e. pipe length, diameters and pipe roughness coefficients, node elevations and daily average consumption volume at each node. In addition, historical flow-data are provided, measured at the supply node, and are assumed to describe the normal operation over all nodes.

Let $m = 198$ and $\mathcal{V} = \{1, 2, ..., 198\}$. A $\Delta t = 5$ minute time step for computing the hydraulic and quality dynamics is considered; each 24-hour period is therefore comprised of $T_h = \frac{24 \cdot 60}{5} = 288$ time steps. There are $N_w = 100$ nodes in the network which outflow water to consumers; the set of demand nodes is $\mathcal{W} = \{w_1, ..., w_{100}\}$, $\mathcal{W} \subset \mathcal{V}$. Only one impact metric is considered ($N_\xi = 1$), the volume of contaminated water consumed at demand node $w_i \in \mathcal{W}$. This is computed using

$$f_\xi(x_{w_i}(k), d_{w_i}(k)) = \begin{cases} d_{w_i}(k)\Delta t & \text{if } x_{w_i}(k) > \epsilon_c \\ 0 & \text{if } x_{w_i}(k) \leq \epsilon_c \end{cases}$$

and $f_\psi(\xi(k)) = \xi_1(k) + ... + \xi_{N_w}(k)$, where ϵ_c is the concentration threshold above which the consumed volume is assumed to be polluted. In this example, we let $\epsilon_c = 0$.

5.2 Framework Parameters

In constructing the contamination scenarios, a fault is modelled as a step-function corresponding to contaminant concentration $1\frac{mg}{L}$, which starts at a certain time within the first time period and is terminated when the fault has been detected and the distribution system has been stopped. A singe, zero stopping time-delay is considered, $\mathcal{T}_d = \{0\}$. We further assume that $\varphi(k) = \Theta p(k)$, such that $\Theta \in \mathbb{R}^{100 \times 48}$ and $p : \mathbb{N}^+ \mapsto \mathbb{R}^{48}$ with $z = 48$ basis unit-step functions. In specific, $p_i(k) = u(k - 1 - 6(i - 1))$, for $i = 1, ..., 48$, where $u(k) = 1$ if $k \geq 0$ and $u(k) = 0$ otherwise. From the set of all fault scenarios $\mathcal{H}^*$, we construct the finite set $\mathcal{H} \subset \mathcal{H}^*$ through grid sampling for scenarios which initiate every half hour; the finite set $\mathcal{H}$ is comprised of $N_h = 4800$ single source contamination scenarios with concentration $1\frac{mg}{L}$.

Furthermore we consider that the candidate sampling nodes set $\mathcal{V}_s$ is the same as the set of demand nodes $\mathcal{W}$, $\mathcal{V}_s \equiv \mathcal{W}$. Thus, we construct the set $\mathcal{Q}_s$ with $N_{qs} = 100$ two-tuples. According to the problem specifications, we construct the finite sampling-times set, $\mathcal{T}_s$. Assuming that manual sampling can be conducted every 30 minutes, then in the case when sampling can be performed "any time" within a day, the set is $\mathcal{T}_s = \{6, 12, ..., 288\}$. In the case of "working hours" constraints, e.g. 8pm-5am, then the set is $\mathcal{T}_s = \{96, 102, ..., 204\}$. In this example, for the "any-time" sampling case, comprised of 48 discrete time instances and 100 candidate sampling node, the set $\mathcal{Q}_m$ is constructed, with $N_{qm} = 4800$ elements. Finally, the set $\mathcal{Q} = \{\mathcal{Q}_s, \mathcal{Q}_m\}$ is comprised of $N_q = 4900$ two-tuple elements.

If a fault is not detected by a sensor or through manual sampling, it is detected by other means $T_{other} = 288$ time steps after it first occurs. For $\lambda = 2$, we further define the discrete maximum time $T_{max} = \lambda T_h = 576$ which will be considered in the simulations, and $\mathcal{T} = \{1, 2, ..., T_{max}\}$ is the set of all time steps considered.

For computing the overall-impact matrix Ω, it is assumed that the time-varying matrix $A(k)$ is known for all times, that no reactions occur in the network (i.e. $\zeta(x(k)) = 0$), that the contaminant detection threshold is zero (i.e. $\epsilon = 0$), and that the outflows $d_w(k)$ for all demand nodes $w \in \mathcal{W}$ are known. By using the advection and impact dynamics, the overall-impact matrix Ω is constructed, with size 4800×4900; in addition, for $N_\xi = 1$, $\mathcal{J} = \{\Omega\}$.

In order to provide more intuition on the information included in Ω, we provide the frequencies of the maximum normalized overall-impact for each scenario, depicted in Fig. 2. The tail distribution shows that a few scenarios can cause extreme damage, and in addition that a large number of scenarios have impacts greater than half the worst-case. To simplify the problem, we might want to neglect scenarios with impact less than a certain percentage of the maximum impact.

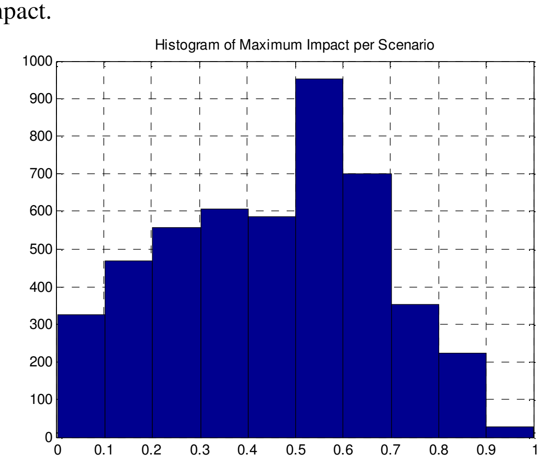

Figure 2. Histogram of the normalized worst-case impact for all scenarios considered

5.3 Reducing Contamination Risk

In order to detect the neuralgic locations which may cause the top-5% of the worst-faults, we compute a set of $\mathcal{Y}_0$ of scenario indices, corresponding to some node locations which should be secured by physical means. We observe that the worst contamination faults originate from a specific location in the network. This is the first step of a complete security and fault-diagnosis scheme.

As discussed, contamination risk is further reduced by installing quality sensors in the network and by conducting manual sampling. To demonstrate the risk reduction when sensors and manual sampling are considered, we follow the solution methodology proposed. We define $\mathcal{L}_s = \{1, ..., 100\}$ and $\mathcal{L}_m = \{101, ..., 4900\}$. In addition, for a certain number of sensors M_s and a certain number of manual sampling M_m, we define the set of all possible solutions $\mathbb{X}$. Let $\mathcal{G} = \{1, ..., 4800\}$ be the set of all scenario indices considered.

We first examine a single-objective problem, so as to compute a solution for various numbers of sensors and manual samplings in the network, using the Average Impact risk-metric; the optimization problem is defined as

$$\mathcal{Y} = \underset{\mathcal{X}\in\mathbb{X}}{\operatorname{argmin}} \{F_1(\mathcal{X};\Omega)\}$$

$$F_1(\mathcal{X};\Omega) = \frac{1}{|\mathcal{G}|}\sum_{i\in\mathcal{G}} \min_{j\in\mathcal{X}} \Omega_{(i,j)}.$$

The results for various cases are summarized in Table 1.

Sensors M_s	Manual M_m	Average Impact
0	0	45.13%
0	1	7.20%
0	2	4.00%
1	0	0.70%
1	2	0.58%
2	0	0.45%
2	2	0.39%
2	4	0.35%
4	0	0.32%

Table 1. Normalized average impacts for various sensor and manual sampling schemes (100% = maximum impact)

We observe that in the case when no sensors or manual sampling is considered, the average impact is 45.13%, where 100% is the maximum impact. The results confirm the intuition that manual sampling will decrease the average impact, but not as much as the placement of sensors in the network. Another observation from the results is that the rate of decrease in the average impact reduces when the number of sensors is increased. Therefore, beyond a certain number of sensors, the decrease of risk may not be significant enough to justify the installation and maintenance costs.

To demonstrate the case where multiple-objectives are considered, we consider the problem of deciding for two manual sampling locations and times, where $M_s = 0$ and $M_m = 2$. The optimization program is formulated as:

$$\begin{aligned}\mathcal{Y} &= \underset{\mathcal{X}\in\mathbb{X}}{\operatorname{argmin}} \{F_1(\mathcal{X};\Omega), F_2(\mathcal{X};\Omega)\}\\ F_A(\mathcal{X};\Omega) &= \frac{1}{|\mathcal{G}|}\sum_{i\in\mathcal{G}} \min_{j\in\mathcal{X}} \Omega_{(i,j)}\\ F_B(\mathcal{X};\Omega) &= \max_{i\in\mathcal{G}} \min_{j\in\mathcal{X}} \Omega_{(i,j)}\end{aligned}$$

and the Pareto front computed is depicted in Figure 3. Each of the 10 elements on the Pareto front corresponds to a set of manual sampling location and times, from which, the decision maker would choose a single solution, using high-level reasoning.

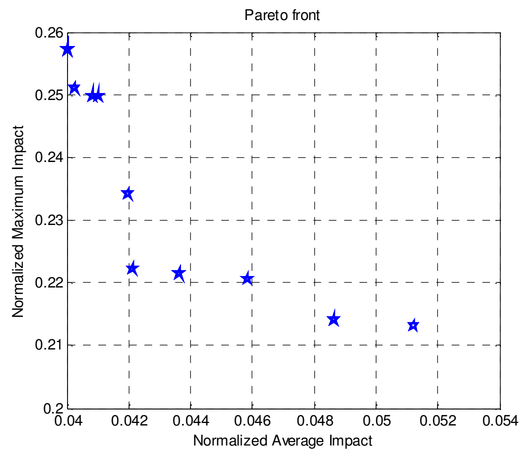

Figure 3. Pareto front of solutions for the multi-objective example.

The optimization problems were solved using the *Matlab Optimization Toolbox*. In specific, for the single-objective problem, the genetic algorithm tool *gatool* was utilized by creating a population of 1 000 solutions and by using rank fitness scaling and stochastic uniform selection. 80% of the population crossovers through a scattered function and the remaining population are subject to mutation through an adaptive function. For the multi-objective case, the genetic algorithm tool *gamultiobj* was utilized by creating a population of 200 solutions. Crossover and mutation parameters are the same as in the single-objective, while the most fit population is kept down to 70%, in order to maintain a diverse population. The implementation is based on a variant of the elitist multi-objective evolutionary algorithm NSGA-II [15].

6. Conclusions and Future Work

Critical Infrastructure security, especially in relation to drinking water distribution networks, is receiving increasing attention. In the present work, we developed a mathematical framework suitable for addressing various security and fault diagnosis problems in water distribution networks. In specific, we examined the static problem of selecting a number of locations to install water quality sensors, as well as locations and times where and when manual sampling should be performed, in order to reduce the damage risk due to a contamination. In the mathematical framework we addressed the problem of computing from a set of contamination scenarios and a set of feasible locations and times in the network, the impact or "damage" caused by each fault, which is used as a metric of the effectiveness of various security schemes. Various security issues were addressed and an optimization solution methodology was proposed, which was demonstrated using a real water distribution network under realistic operational conditions.

Protecting the consumers through quality sensor placement and manual sampling constitutes the first step towards a complete security scheme which would include more advanced techniques, such as using the water quality dynamics for fault detection, as well as adaptively change the manual sampling decisions in order to minimize the contamination risk, if the system parameters change in time.

Acknowledgment

This work is partially supported by the Research Promotion Foundation (Cyprus).

References

[1] T. M. Walski and J. W. Male, "Maintenance and rehabilitation /replacement," in *Water distribution systems handbook*, L. Mays (ed.) New York: McGraw-Hill, 2000, pp. 17.1–17.28.

[2] C. Toregas and C. ReVelle, "Optimal location under time or distance constraints," *Papers in Regional Science*, 28(1), pp. 131–143, Dec. 1972.

[3] R. Church and C. ReVelle, "The maximal covering location problem," *Papers in Regional Science*, 32(1), pp. 101–118, Dec. 1974.

[4] B. Lee and R. Deininger, "Optimal locations of monitoring stations in water distribution system," *ASCE Journal of Environmental Engineering*, 118(1), pp. 4–16, Jan./Feb. 1992.

[5] S. Hakimi, "Optimum locations of switching centers and the absolute centers and medians of a graph," *Operations Research*, 12, pp. 450–459, May/Jun. 1964.

[6] J. Berry, L. Fleischer, W. Hart, and C. Phillips, "Sensor placement in municipal water networks," in *Proceedings of the ASCE World Water and Environmental Resources Congress*, 2003, pp. 40–49.

[7] J. Berry, W. Hart, C. Phillips, J. Uber, and J. Watson, "Sensor placement in municipal water networks with temporal integer programming models," *ASCE Journal of Water Resources Planning and Management*, 132, pp. 218–224, Jul./Aug. 2006.

[8] J.-P. Watson, H. J. Greenberg, and W. E. Hart, "A multiple-objective analysis of sensor placement optimization in water networks," in *Proceedings of the ASCE World Water and Environmental Resources Congress*, 2004, pp. 456–465.

[9] A. Ostfeld, J. G. Uber, E. Salomons, J. W. Berry, W. E. Hart, C. A. Phillips, J.-P. Watson, G. Dorini, P. Jonkergouw, Z. Kapelan, F. di Pierro, S.-T. Khu, D. Savic, D. Eliades, M. Polycarpou, S. R. Ghimire, B. D. Barkdoll, R. Gueli, J. J. Huang, E. A. McBean, W. James, A. Krause, J. Leskovec, S. Isovitsch, J. Xu, C. Guestrin, J. VanBriesen, M. Small, P. Fischbeck, A. Preis, M. Propato, O. Piller, G. B. Trachtman, Z. Y. Wu, and T. Walski, "The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms," *ASCE Journal of Water Resources Planning and Management*, 134(6), pp. 556–568, Nov./Dec. 2008.

[10] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, and C. Faloutsos, "Efficient sensor placement optimization for securing large water distribution networks," *ASCE Journal of Water Resources Planning and Management*, 134(6), pp. 516–526, Nov./Dec. 2008.

[11] R. LeVeque, "Nonlinear conservation laws and finite volume methods," in *Computational methods for astrophysical fluid flow*, O. Steiner and A. Gautschy, (eds.), Berlin, Germany: Springer, 1998.

[12] D. Rosen and S. Zenios, "Enterprise-wide asset and liability management: Issues, institutions and models," in *Handbook of Asset and Liability Management*, S. Zenios and W. Ziemba, (eds.), North-Holland, 2006.

[13] S. Uryasev, *Probabilistic Constrained Optimization: Methodology and Applications*. Springer, 2000.

[14] S. Rao, *Engineering Optimization: Theory and Practice*. Wiley-Interscience, 1996.

[15] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, 6(2), pp. 182–197, Apr. 2002.

Author Biographies

Demetrios G. Eliades received the Diploma in electrical and computer engineering from the Aristotle University of Thessaloniki, Greece, in 2004, the M.Sc. degree in artificial intelligence from the University of Edinburgh, Scotland, in 2005, and is currently a Ph.D. candidate in electrical engineering at the University of Cyprus, Nicosia, Cyprus. His research interests include critical infrastructure security, optimal location problems, and computational intelligence.

Marios M. Polycarpou is a Professor of Electrical and Computer Engineering and Director of the KIOS Research Center for Intelligent Systems and Networks at the University of Cyprus. He received the B.A. degree in Computer Science and the B.Sc. degree in Electrical Engineering both from Rice University, Houston, TX, USA in 1987, and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Southern California, Los Angeles, CA, in 1989 and 1992 respectively. In 1992, he joined the University of Cincinnati, Ohio, USA, where he reached the rank of Professor of Electrical and Computer Engineering and Computer Science. In 2001, he was the first faculty member and the founding department Chair of the newly established Electrical and Computer Engineering Dept. at the University of Cyprus. His teaching and research interests are in intelligent systems and control, adaptive and cooperative control systems, computational intelligence, fault diagnosis and distributed agents. Dr. Polycarpou has published more than 185 articles in refereed journals, edited books and refereed conference proceedings, and co-authored the book Adaptive Approximation Based Control, published by Wiley in 2006. He is also the holder of 3 patents.

Prof. Polycarpou is currently the Editor-in-Chief of the IEEE Transactions on Neural Networks. He serves as an Associate Editor of two international journals and past Associate Editor of the IEEE Transactions on Neural Networks (1998-2003) and of the IEEE Transactions on Automatic Control (1999-2002). He served as the Chair of the Technical Committee on Intelligent Control, IEEE Control Systems Society (2003-05) and as Vice President, Conferences, of the IEEE Computational Intelligence Society (2002-03). He is currently an elected member of the Board of Governors of the IEEE Control Systems Society and an elected AdCom member of the IEEE Computational Intelligence Society. Dr. Polycarpou was the recipient of the William H. Middendorf Research Excellence Award at the University of Cincinnati (1997) and was nominated by students for the Professor of the Year award (1996). His research has been funded by several agencies in the United States, the European Commission and the Research Promotion Foundation of Cyprus. Dr. Polycarpou is a Fellow of the IEEE.

Bambos Charalambous holds a BSc degree in Civil Engineering and a Masters in Business Administration. He is currently a PhD candidate working on 'Sustainability of water use in the Mediterranean with particular application to Cyprus'. He is a chartered engineer (CEng) registered with the UK Engineering Council, a corporate member of the Institution of Civil Engineers (UK) and a

Fellow of the Institution of Water and Environmental Management (UK). His experience in water related areas spans over 30 years and has worked on many projects in Europe, West Africa, Asia, United Arab Emirates and Cyprus. Since 1995 he is heading the Technical Services of the Water Board of Lemesos, Cyprus, with responsibilities over all technical matters of the Board, including planning, operation and maintenance of the water supply system, as well as the training and continuous professional development of all the staff of the Technical Services. He has instigated and is overseeing the implementation of the Water Board's Water Loss Management strategy. He is former Chair of the International Water Association's Water Loss Task Force and present Chair of the Water and Energy Task Force. He has published numerous papers on issues related to the management of urban water supply distribution systems.