International Conference on Hydroinformatics

2014

Lessons Learned In Solving The Contaminant Source Identification In An Online Context

Hervé Ung

Olivier A. Piller

Deuerlein Jochen

Denis Gilbert

Montalvo Idel

LESSONS LEARNED IN SOLVING THE CONTAMINANT SOURCE IDENTIFICATION PROBLEM IN AN ONLINE CONTEXT

UNG HERVÉ (1), PILLER OLIVIER (1), DENIS GILBERT (1), DEUERLEIN JOCHEN (2), MONTALVO IDEL (2)

(1): Irstea, 50, avenue de Verdun, F-33612 Cestas (France)

(2): 3S Consult GmbH, Office Karlsruhe, Albtalstrasse 13, D-76137 Karlsruhe, Germany

Protection of Water Distribution Networks (WDNs) against contamination events is of paramount importance. Either deliberate or accidental contamination of this infrastructure has strong negative consequences from both social and economical points of view. The project *SMaRT-Online*^{WDN} aims to develop methods and software solutions to 1) detect contamination from non-specific sensors, 2) maintain an online water quantity and water quality model that is reliable and 3) use the past model predictions to backtrack the potential sources of contamination. The problem of source identification consists of determining the location and duration of a contamination taking into account sensor responses. Our solution is a two-step enumeration/exploration method. Firstly, we solve the transport equation in reverse time for enumeration of the potential solutions. This is made independent of the reaction kinetics of particular substances. The known boundary conditions are the responses of sensors that count successive contaminant fronts arriving at each sensor. In the second exploration step a probability calculation for ranking of the candidate solutions is proposed with two general stochastic methods (minimum relative entropy or least squares methods). An extensive use of simplification methods is carried out, both temporally and spatially on the dynamic graph. A sensitivity analysis is made with regards to the demand uncertainty. Results on real networks in France and Germany are presented.

INTRODUCTION

The source identification problem has been treated by a large number of authors. Their solutions can be classified in three categories. The first type consists of the enumeration of all possible sources of contamination: the solution is generally a backtracking algorithm as described in the paper of Shang *et al.* [1] with particle tracking. The second category explores the potential contaminations by giving them weights (*e.g.*, use of Minimum Relative Entropy formulation in [2]). Finally the third category consists of taking into account non-perfect sensors. Several authors [3] used Bayesian Belief Networks with non-perfect sensors and other information such as operation changes. This paper will follow the work of Propato *et al.* [2] as well as Ung *et al.* [4] and adapt it to real time monitoring. It describes the methods used and the different difficulties that have been met, and their solutions.

Enumeration exploration of the potential sources

Here two types of solution are proposed. The first one is an enumeration/exploration solution as in [4] and the second uses the adjoint equation.

The transport equation is frequently used in hydraulics software to estimate the quality inside the pipes, for instance for the transport of a conservative chemical contaminant in the network.

$$\frac{\partial C}{\partial t} + U(t)\frac{\partial C}{\partial x} = 0, \qquad (1)$$

Where C is the concentration of the contaminant.

Another way to resolve the source identification problem is by deriving the adjoint equations [5][6].

$$\frac{\partial \psi}{\partial t} - U(t)\frac{\partial \psi}{\partial x} = 0, \qquad (2)$$

Where τ is backward time, and Ψ is the adjoint state, here the sensitivity of concentration with regards to the mass introduced. The advantage of using the adjoint model Eq. (2) compared to using Eq. (1) is that only one resolution of (2) is needed instead of multiple of (1) for each node. Solved with appropriate boundary conditions and Initial Conditions, the adjoint equation gives the adjoint state temporal pattern at every node.

The other two-step method consists of a step of enumeration and a step of weighting or exploration. The first step lists the different possible nodes of contamination that can explain the sensor responses. It also uses negative answers to take away nodes from the list as described in Figure 1. A kind of particle backtracking algorithm is used, the solutions are the red nodes, green nodes are not possible and black nodes are unknown.

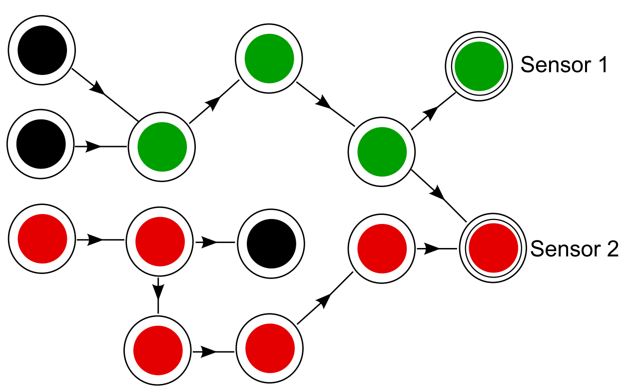

Figure 1: Example of use of the backtracking algorithm

Then an Input/Output matrix is built [4] that explain the relation between the possible source of contamination and the sensors. This matrix will be used in the second step to determine the probability of each node to be the source.

Two stochastic optimisation methods have been used. The first one is the Minimum Relative Entropy method defined by Shanon (1948) that minimized the Kullback-Leibler divergence:

$$\min \mathrm{H}_e(q) := \int_{\Omega} q(\mathbf{U}) \ln\left(\frac{q(\mathbf{U})}{p(\mathbf{U})}\right) d\mathbf{U}$$

subject to:

$$\int_{\Omega} q(\mathbf{U}) d\mathbf{U} = 1 \tag{3}$$

$$\mathbf{A}\,\mathrm{E}_{\mathbf{U}}(q) = \mathbf{1}_m, \quad \mathrm{E}_{\mathbf{U}}(q) = \int_{\Omega} q(\mathbf{U}) \mathbf{U} d\mathbf{U}$$

Where He is the objective to minimize; U is the unknown contaminant intensity vector with Ui between 0 and 1; U is a random variable with joint probality density function (pdf) q; q is to be determined and p is a prior pdf, and EU(q) is the expectation of U with joint pdf q.

The problem is resolved by a Lagrangian approach and we get:

$$\hat{q}(\mathbf{U}, \mu, \lambda) = p(\mathbf{U}) \exp(-1 - \mu - \lambda^T \mathbf{A}\mathbf{U}) \tag{4}$$

With μ is the Lagrange multiplier of the *q* pdf integrity constraint; and $\boldsymbol{\lambda}$ is the Lagrange multiplier vector associated with the I/O matrix constraint (the information from the sensor responses).

REAL TIME INTEGRATION

In the *SMaRT-Online*^{WDN} project the source identification algorithm is connected to a SCADA system via an OPC-client that was developed by 3S Consult to get data in real time. The methods need an adaptation to use this system in its formulation. For online application of the proposed method the flow velocities of the last few hours have to be known. There exists a trade-off between calculation time, memory consumption and accuracy due to the following reasons:

The most efficient approach in terms of calculation time would be to store the calculated flow velocities for all calculated time steps and all pipes. However, the large amount of data prevents them being saved in memory, so that the flow vectors have to be stored in a database losing further efficiency. An alternative is to store only selected time steps either by constant time steps or dependent on the rate of change. It is easy to see that this approach results in loss of information and accuracy.

The third alternative is to recalculate the hydraulic state of the system for the last few hours (reconstruction calculation). At this stage, it cannot be definitely decided which of the described methods is the most efficient.

With the existing software (*Sir OPC* Drive), reconstruction calculations are already possible and can be used without any additional changes. It is also planned to implement the database version where the pipe velocity vectors of each time step are stored in a kind of FIFO queue see Figure 2. This means that an array of flow vectors over time (size Δt) is created. The array has fixed size and for every new flow vector that is calculated the flow vector that refers to t -Δt is removed from the vector queue.

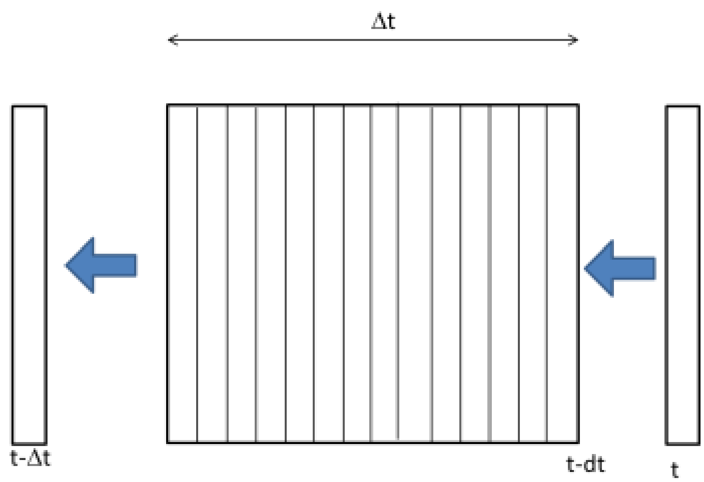

Figure 2: Sliding windows concept for pipe velocity storage in a FIFO queue

SIMPLIFICATION

Finally to tackle large-sized networks, simplification is used. There are two types of simplification - in space and in time. The first type consists of simplifying the graph of the network. The concept is explained in Figure 3. The forest is removed from the original graph (core Figure 3b) then the intermediate nodes of degree equal to 2, to finally obtain the simplified graph that is a topological minor of the original graph (same structure) [7,8].

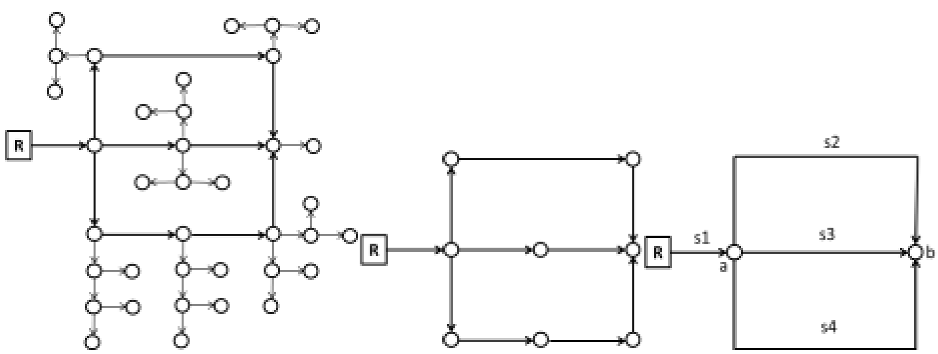

Figure 3. Graph simplification: a) original graph, b) core, c) topological minor

Another kind of simplification consists of aggregating multiple time steps together, for instance put together consecutive times at a node that can explain also consecutives times of positive responses at a sensor. The direct consequence is the reduction of the size of the Input/Output matrix, and therefore a faster calculation.

CONCLUSION

This research aims to adapt existing backtracking methods to online monitoring and security management. Two backtracking methods are compared. The adjoint method is a deterministic one that solves the adjoint equations in reverse time. The enumeration/exploration method is a stochastic one that backtracks particles from binary sensor responses to build an Input/Output matrix first and enumerate all the potential solutions. Then, the potential solutions are ranked by use of a minimum relative entropy method that calculates their expectations and the main quantiles.
This method has been adapted to real time monitoring through the technology of *Sir3S OPC* server to retrieve the calibrated velocities from the recent historical database. Finally, to accelerate the method, simplification methods have been evaluated.

Acknowledgements
The project is supported by the German Federal Ministry of Education and Research (BMBF; project: 13N12180) and by the French Agence Nationale de la Recherche (ANR; project: ANR-11-SECU-006).

REFERENCES

[1] Shang, F., Uber, J. G. & Polycarpou, M. M., 2002. Practical back-tracking algorithm for water distribution system analysis. Journal of Environmental Engineering, 128(5), pp. 441 - 450.
[2] Propato, M., Tryby, M. E. & Piller, O., 2007. Linear Algebra Analysis for contaminat source identification in water distribution systems. Tampa, Florida, ASCE, World Environmental Water Resources Congress, pp. 1 -10.
[3] Perelman, L., and Ostfeld, A. (2013). "Bayesian Networks for Source Intrusion Detection." Journal of Water Resources Planning and Management, 139(4), 426-432.
[4] Ung, H., Piller, O., Gilbert, D. & Mortazavi, I., 2013. Inverse Transport Method for Determination of Potential Contamination Sources with a Stochastic Framework. World Environmental Water Resources Congress 2013: Showcasing the Future, ASCE, pp. 798 - 812. Ung et al.
[5] Neupauer, R. (2011). "Adjoint Sensitivity Analysis of Contaminant Concentrations in Water Distribution Systems." Journal of Engineering Mechanics, 137(1), 31-39.
[6] Wagner (2013). "Source Identification in Water Distribution Systems using the Adjoint Method with Non-Ideal Sensors and Non-Detect Measurements". Environmental Water Resources Congress 2013: Showcasing the Future, ASCE, pp. 975 - 981.
[7] Deuerlein, J. W.(2008): Decomposition model of a general water supply network graph. ASCE Journal of Hydraulic Engineering, 134, 6, 2008, 822 - 832.
[8] Deuerlein, J.; Simpson, A. R. (2012): "A modified global gradient algorithm based on topological decomposition" 14th Annual Symposium on Water Distribution Systems Analysis, WDSA 2012, ASCE, Adelaide, Australia, 24-27 September, 2012.