

The human sensor: real time allocation of leaks and contaminations using tweets and complaints data

G.I.M. Worm^{1/2}, T. Lapikas³, L.C. Rietveld²

¹ PWN, PO Box 2113, 1990 AC, Velsbroek, the Netherlands (e-mail: ignaz.worm@pwn.nl)

² Delft University of Technology, Faculty of Civil Engineering and Geosciences, Department of Water Management, PO Box 5048, 2600 GA, Delft, the Netherlands (e-mail: l.c.rietveld@tudelft.nl)

³ UReason, Pompoenweg 9, 2321 DK, Leiden, the Netherlands (e-mail: tlapikas@ureason.com)

Keywords

Crowd sourcing, back tracing, network model, tweets, complaints

INTRODUCTION

The allocation of a leak or contamination in the distribution network can take days, as can be concluded from the Rijsenhout case (the Netherlands), where a farmer pumped collected rainwater in the distribution network via an illegal connection. Eight days after the first complaints were registered by the water supply company, the location of the contamination was determined. Extensive research efforts are made worldwide to develop sensors to measure deviances of the water quality (Lambeck, 2006) and to monitor the hydraulics and water quality in distribution networks (Allen *et al.*, 2011). Placing sensors and transporting the data is expensive, since the required number of sensors is large and maintenance and security require significant efforts. These high costs give reason to investigate optimal sensor placement in water distribution networks (Berry *et al.*, 2005; Krause *et al.*, 2008; Shastri & Diwekar, 2006).

During a more recent calamity, the shut down of pumping station Hoofddorp as a consequence of a power supply interruption on December 27th 2010, a customer created a map of the affected region, showing the tweets (Twitter messages) related to the interruption of the water supply and published it online, within one hour of the shut down. It appeared that these map and the tweets of an employee of the water supply company played an important role in informing the customers.

A method is known to locate a contamination source using a network model focusing on flow directions (Davidson *et al.*, 2005). Another method combines and analyses internally available data and uses EPANET calculation results to determine where a water particle (possibly) has come from to identify the source of a measured water quality deviation (Besner *et al.*, 2005). Back tracing in looped distribution networks is more complicated than in branched networks, especially when taking into account that flow directions might change as a consequence of changes in demands.

This research aims to shorten the time from the moment a contamination or leak occurs in a drinking water distribution network to the moment of the allocation of this disturbance by feeding customer's responses to a back tracing network model. The back tracing distribution network model will be fed with i) tweets concerning the disturbance, ii) complaint data logged by the water supply company and iii) missed telephone calls by the water supply company. The hypothesis is that the customers of a water supply company are rapid, reliable, cheap and dense sensors for disturbances of the drinking water quality and delivery. The intended end-users of the system are customers visiting the company's website during calamities, operators and field maintenance personnel.

Is it ethically responsible to 'use' customers as a sensor for water quality and delivery? Using customers as sensors is as old as the modern drinking water supply in Europe. In 1849 John Snow (Snow, 1855) analysed the number and location of casualties as a consequence of a cholera outbreak in London. With the map he composed he was able to determine the location of the contaminated well, see Figure 1. The distinctive point is that the customers are not 'used' as sensor,

but information is created from data they provide voluntarily to social media or to the water supply company. Customers are expected to appreciate that the information they supply is taken into account and used to solve the problem and to supply information.



Figure 1. Casualties in London in 1849 as a consequence of cholera infection, based on (Snow, 1855).

MATERIALS AND METHOD

Materials

An EPANET (Rossman, 2000) distribution network will be equipped with an existing or new back trace algorithm. The main challenge is to determine the path water followed in a looped network. The input for the model (tweets, complaints data and missed telephone calls) are filtered, converted into a uniform format and collected in a database. The graphical output of the model is the estimated location of the source of the disturbance on a map, encircled by a bandwidth of uncertainty, which shrinks when more input enters the model. An algorithm is defined to determine the uncertainty of the results stochastically, based on the density (and possibly other characteristics of the input data). An existing application programming interface (API) is used to extract relevant information from Twitter. Interface will be set up to export customer complaints data from the SAP/Customer Relations Management (CRM) system of water supply company PWN and missed calls from the telephone exchange.

A generic system will be developed to set up the human sensor for new customers (i.e. water supply companies) efficiently. The system is initially able to set up the interfaces with Twitter, the CRM package of the company and the telephone exchange, and will be easily extendible for new data sources like complaints websites or other CRM packages. The generic system will provide a graphical presentation of this data on a map and feed the data to the back tracing network model.

Use cases will be defined and a scheme of objectives. The system will support flexible and light-weight data-integration techniques.

Method

With the generic system and the EPANET platform with backtrace function, a human sensor will be set up for a district metered area of PWN. To test the effectiveness of the sensor, a virtual contamination will be made in the area. Tweets and complaints will be created in a way experts suggest they would occur during such a contamination. The tweets and complaints are processed as real tweets and complaints and the response of the human sensor will be assessed using predetermined criteria.

RESULTS AND CONCLUSIONS

This research is launched at this moment.

AWARD

This idea has been granted the 'Best IT-idea in the water sector' award of the Koninklijk Nederlands Waternetwerk in May 2012.

REFERENCES

- Allen, M., Preis, A., Iqbal, M., Srirangarajan, S., Beng Lim, H., Girod, L. and Whittle, A. J. (2011). Real-time in-network distribution system monitoring to improve operational efficiency. *J. AWWA*, **103**(7), 63-75.
- Berry, J., Fleischer, L., Hart, W., Phillips, C. and Watson, J. (2005). Sensor Placement in Municipal Water Networks. *J. Water Resour. Plann. Manage*, **131**(3), 237-243.
- Besner, M. C., Gauthier, V., Trépanier, M., Leclair, M. and Prévost, M. (2005). Interactive analyser for understanding water quality problems in distribution systems. *Urban Wat. J.*, **2**(2), 93-105.
- Davidson, J., Bouchart, F., Cavill, S. and Jowitt, P. (2005). Real-Time Connectivity Modeling of Water Distribution Networks to Predict Contamination Spread. *J. of Comp. in Civ. Eng.*, **19**(4), 377-386.
- Krause, A., Leskovec, J., Guestrin, C., Van Briesen, J. and Faloutsos, C. (2008). Efficient Sensor Placement Optimization for Securing Large Water Distribution Networks. *J. Water Resour. Plann. Manage*, **134**(6), 516-526.
- Lambeck, P. V. (2006). Integrated optical sensors for the chemical domain *Meas. Sc. and Techn.*, **17**(8).
- Rossman, L. A. (2000). EPANET 2 Users manual. Cincinnati, National Risk Management Research Laboratory, Water supply and water resources division.
- Shastri, Y. and Diwekar, U. (2006). Sensor Placement in Water Networks: A Stochastic Programming Approach. *J. Water Resour. Plann. Manage*, **132**(3), 192-203.
- Snow, J. (1855). Mode of communication of cholera. *MD Brit. and For. Med.-Chir. Rev.*, **15**, 449-463.