Morphing of Water Networks

Jonathan Berry[*] William Hart[†] Carl Laird[‡] James Uber[§]

January 15, 2007

Abstract

Along with increases in the level of research activity surrounding sensor placement and event detection algorithms come greater needs for realistic datasets. In particular, realistic water network models are hard to obtain. Utilities are justifiably reluctant to release their models because of security concerns. The result is that all too many papers have presented algorithmic work evaluated only on networks with fewer than 100 nodes.

In this paper, we explore the idea of "morphing" a water distribution network in order to defeat state-of-the-art layout reconstruction methods. Specifically, we solve the problem of finding new lengths and diameters for the pipes that are maximally different from the original parameters, subject to the constraint that travel times through the network are approximately preserved. The resulting networks are appropriate for evaluating sensor placement algorithms, but otherwise have strange characteristics. For example, some pipes may be extremely long or short, and the pressures and heads may be quite confusing to an adversary attempting to identify the network. We give results for small networks and discuss the challenges of scaling our algorithm up to larger networks.

1 Introduction

Most published research on water distribution system sensor placement and source inversion problems has focused on small networks such as the EPANET example datasets. It is hard to judge the applicability of such work to real networks. It would be beneficial to the water security community if large, realistic datasets were in the public domain and available for experimentation. However, such datasets are proprietary information of water utilities, and releasing them could have national security consequences.

In this paper we present what is, to our knowledge, the first attempt to produce *morphed* water distribution networks. These networks are based on real networks, but are "morphed" in such a way to hide the identity of the original, while approximately maintaining properties useful to researchers studying sensor placement and

[*]Sandia National Laboratories
[†]Sandia National Laboratories
[‡]University of Pittsburgh
[§]University of Cincinnati

event detection. Our morphing model changes the lengths and diameters of pipes, while approximately preserving the travel time of water packets along the pipes. The resulting network confounds the state of the art multidimensional scaling algorithms that an adversary might use in an attempt to lay out the nodes and identify the original network.

This work does not change the topology of the original network, and thus is probably not yet sufficient to produce safely sharable networks with no security risk. Rather, it represents a first step in this direction.

2 Background

The "coordinates" section in an EPANET [7] input file is not used in hydraulic or water quality computations. In many cases, these original coordinates clearly identify the city, as node locations are constrained by geography. For example, the course of a river flowing through a city is traced out by the nearby water disribution system nodes, and often the shape of the city itself is clear. Any secure network data derived from a water distribution model clearly would not contain the original coordinates. Furthermore, simply removing the "coordinates" section or including obfuscated coordinates is not sufficient. There is free software based on a technique called *multi-dimensional scaling* that uses optimization to compute coordinates for the nodes of a network, given a set of *lengths* of the edges in the network. For example, *XGvis* [2] is a multi-dimensional scaling package that is accessible from the "GGobi" visualization software. of Swayne, Cook, and Buja [8].

As we will see in Section 4, XGvis does a credible job of reproducing the network layouts of some small networks, even using its default settings. There are many XGvis parameters that can be varied, and we assume that a knowledgeable adversary would be able to use tools like this to identify the city of any water distribution network that leaked into the public domain with the original pipe lengths intact.

3 A Morphing Model

Sensor placement researchers use both hydraulic and water quality simulations to generate data for their optimization models. The hydraulic characteristics of a network determine flows and velocities of water throughout the network during several discrete flow patterns, then water quality simulations use this information to simulate the propagation of water. Given a source of contamination, these simulations track the concentration of contaminant as it travels.

3.1 Modeling Assumptions

Sensor placement optimization models require data concerning the impact of contamination in a network over large sets of contamination scenarios, or *events*. These data are gathered by running many water quality simulations, each of which as a defined source (or sources) of contamination, an injection strength, and an injection duration. For this paper, we assume that the travel times of contaminated water packets through pipes determine, for each water quality time step, what concentration of contamination arrives at a node. For a given path between a contamination source and a node, we assume that if travel times can be preserved during the morphing process, then the amount of contaminant reaching that node from the source at each time step will be preserved as well.

Thus, our goal is to preserve travel times through the network during the morphing process. Furthermore, since we are interested only in contaminant transport during water quality simulations, we are willing to sacrifice hydraulic intuition in achieving this goal. For example, it is acceptable for the morphed network to have negative pressures or heads. In order to obfuscate the network, it is acceptable, and even desirable to have pipe dimensions that are quite unusual.

3.2 The Formulation

The Darcy-Weisbach formula relates head loss along a pipe (h_e) to flow along that edge (q_e) as follows:

$$h_e = p_e \frac{l_e}{d_e^5} q_e^2.$$

Our strategy is to simultaneously vary the lengths and diameters of all pipes in order to maintain this relationship between head loss and flow at each. Assuming that the hydraulic simulation of a network model is based upon the Darcy-Weisbach formula, our morphing procedure is:

- Find a *spanning tree* of the water network. This is a set of connections (pipes, valves, pumps) such that there is exactly one path from any vertex (junction, reservoir, tank) to any other vertex.
- Find the *fundamental cycles* of that spanning tree. Each non-tree connection forms a fundamental cycle with the unique path in the spanning tree that connects its endpoints.
- Change the lengths and diameters of the pipes in order to make all of the fundamental cycles as nearly equal in length as possible, while enforcing flow and head loss conservation, and while approximately maintaining the travel times along pipes during water quality simulations.

The intuition behind our model is that some fundamental cycles are naturally quite long, while others are very localized. Our model upsets this natural situation by forcing all cycles to be approximately the same length. The result is a new set of pipe lengths and diameters that are radically different from the originals, yet the preserved Darcy-Weisbach relationships between head loss and flow approximately preserves water travel times.

This process is implemented by a nonlinear optimization model, which uses the following notation:

- *edge*: a connection (a pipe, valve, or pump)
- $\mathcal{V}$: the set of all nodes in the network
- $\mathcal{R}$: the set of tanks and reservoirs in the network
- $\mathcal{P}$: the set of pipes in the network
- l'_e: the original length of edge e.

- d'_e: the original diameter of edge e.
- q'_e: the original flow along edge e.
- h'_e: the original head loss along edge e.
- h'_v: the original head at node v.
- τ'_e: the original travel time along edge e.
- l_e: the length of edge e.
- d_e: the diameter of edge e.
- q_e: the flow along edge e.
- h_e: the head loss along edge e.
- h_v: the head at node v.
- τ_e: the travel time along edge e.
- B: the set of edges not in the spanning tree
- C_b: the set of edge in the fundamental cycle formed by $b \in B$.
- $N(v)$: the neighbors of node v
- d_v: the demand at node v.
- p_e: the Darcy-Weisbach proportionality constant for pipe e.
- M_C: the maximum length of a fundamental cycle: $\max_{b \in B}\{l(C_b)\}$
- $P_{s,t}$: the path from s to t in the spanning tree.
- a: $\frac{\pi}{4}$

Given these definitions, the model is as follows:

$$\text{(MORPH)} \quad \text{minimize} \quad \sum_{b \in B} (M_C - l(C_b))^2$$

$$\text{where} \quad \begin{cases} d_v + \sum_{e \in N(v)} q_e = 0 & \forall v \in \mathcal{V} \\ \sum_{e \in C_b} h_e = 0 & \forall b \in B \\ \sum_{e \in C_b} h_e = 0 & \forall b \in B \\ h_e = p_e \frac{l_e}{d_e^5} q_e^2 & \forall e \in \mathcal{P} \\ \frac{q_{(i,j)}}{\sum_{(x,j) \in \mathcal{P}} q_{(x,j)}} = \frac{q'_{(i,j)}}{\sum_{(x,j) \in \mathcal{P}} q'_{(x,j)}} & \forall e = (i,j) \in \mathcal{P} \\ \tau_e = \frac{a l_e d_e^2}{q_e} & \forall e \in \mathcal{P} \\ \tau_e = \tau'_e & \forall e \in \mathcal{P} \\ \sum_{e \in P_{s,t}} h_e = h_s - h_t & \forall s,t \in R \end{cases}$$

The first set of constraints enforces flow conservation; the second set enforces energy conservation (head loss around each cycle is zero); the third set enforces the Darcy-Weisbach formula; the fourth conserves the flow ratio (flow of pipe e with respect to total flow into e's destination). The fifth set of constraints defines travel time along edge e, and the sixth preserves travel times. Finally, the seventh set of constraints preserves head loss between water sources (tanks and reservoirs). Note that the equalities in this model are implemented as near-equalities, with tolerances.

The actual implementation was slightly more complicated than the MORPH model above. For example, pump curves coefficients were modeled, the flows and head losses of valves were held invariant, the flows along each connection were broken down into direction and magnitude, and there were bounds on the extent to which lengths, diameters, and flows could change. We also included a constraint that maintained the original head loss along paths connecting water sources such as reservoirs.

The model was tested on small EPANET example networks and a suite of simple rectangular grid networks. We used AMPL [4] to compute the spanning tree and fundamental cycles of the input network, and to construct the optimization problem. AMPL invoked the IPOPT [9] nonlinear solver to compute new pipe lengths and diameters.

Since the Darcy-Weisbach formula is highly non-linear, the optimization problem is difficult to solve, and IPOPT often would not converge to a solution for inputs with thousands of nodes. However, we have a reformulation of MORPH on which IPOPT may perform better. That reformulation is not reported in this paper, though. In Section 4, we report our preliminary results on networks of roughly 100 nodes.

3.3 Datasets

We present results for two different datasets: a 10×10 grid network, and the 97 node "EPANET example 3." The former consists of 101 nodes: a single reservoir at an elevation of 1500 feet, and 100 junctions. The latter are numbered in row-major order (English reading order) and have monotonically decreasing elevations. These start from 1000 feet, and decreasing by half a foot per vertex. The reservoir is the single source of injection, and demands are randomly distributed among the 100 grid vertices.

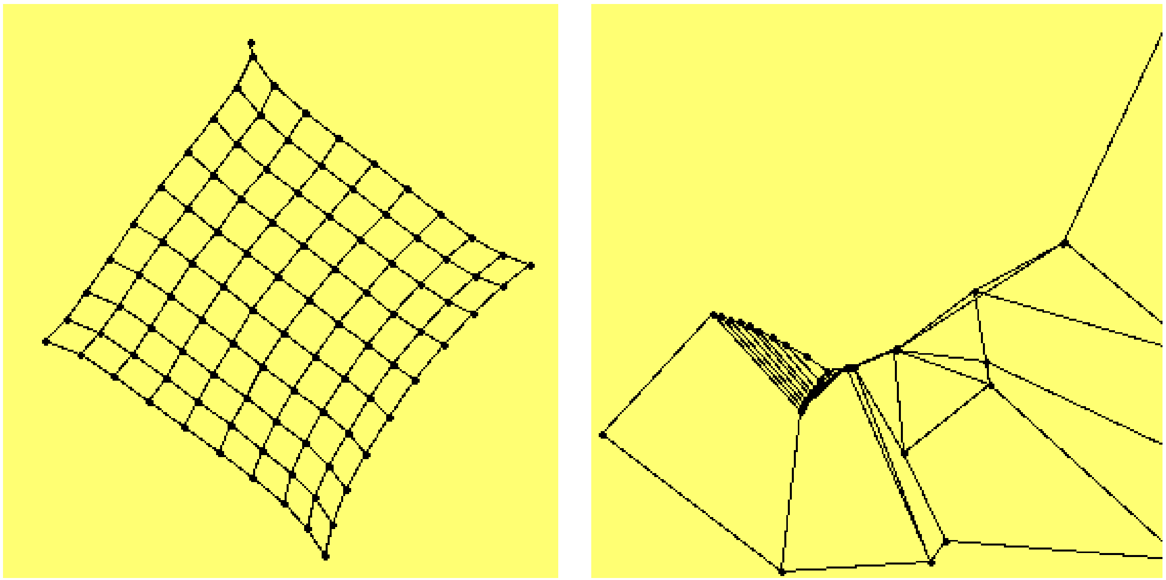
Our second dataset is the 97 node "EPANET example 3." This network has 92 nodes, two reservoirs, three tanks, 117 pipes, 2 pumps, and no valves. There are 59 nodes with non-zero demands.

4 Results

We carried out the morphing process on each of the two datasets, then explored the ability of multidimensional scaling software to reproduce the network drawing using only pipe length information. These visual results are presented in Section 4.1 below. Next, we ran sensor placement algorithms from [1] on both the original and morphed versions of the networks. These results are presented in Section 4.2 below.

4.1 Visual Results

Consider Figures 1 and 2. The (a) pane of each figure shows the result of applying XGvis multidimensional scaling (with default settings) to the original network and visualizing the result using GGobi. In each case, the result is good enough to identify the original network. The (b) panes of these figures show the result of applying the identical tools to networks with new pipe lengths and diameters determined by our MORPH model. In both cases, the new lengths confound multidimensional scaling and the original network layout remains hidden.



(a) (b)

Figure 1: A simple grid network drawn with coordinates obtained from multi-dimensional scaling with (a) the original pipe lengths, and (b) the morphed pipe lengths

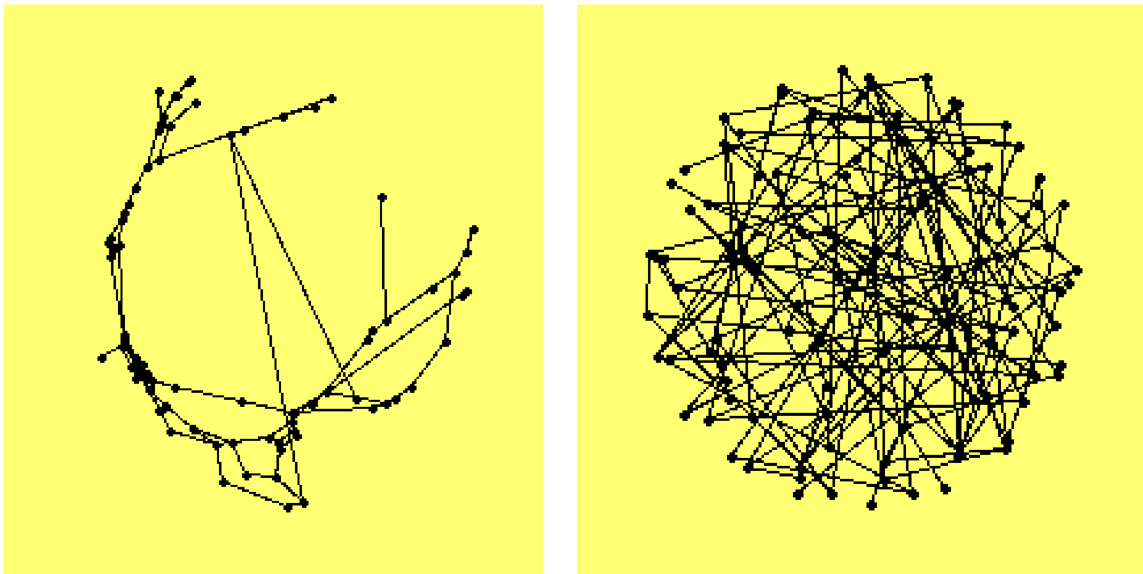
4.2 Computational Results

The motivation for morphing is to provide datasets that do not betray their location, yet behave much like the originals during water quality simulations. We compared the performance of our original and morphed networks using the SPOT [5], software tool for running large ensembles of water quality simulations and using the results to find sensor layouts that minimize the impact of contamination events.

For the both networks, we generated contamination event ensembles as follows. For each of four different starting times throughout the day (3am, 9am, 3pm, and 9pm), we ran 59 water quality simulations – one for each non-zero demand node. The grid network had only one flow pattern: gravity flow from the high ground to the low ground. The EPANET example had four flow patterns. We used TEVA [**?**] to control the runs of our water quality simulations, and SPOT [5] to generate files containing sequences of node hits (contamination events) and the assocatied network-wide impact at the moment of those hits. After computing the impacts, we ran the PICO [3] solver that is included with SPOT to find optimal sensor placements that minimize impact across all contamination events. SPOT allows the user to select many variations of this objective, and we chose to minimize the expected impact (as opposed to the worst case, etc.).

We explored two different types of impact values: *contaminant mass consumed* in generic units (e.g. cells for a biological contaminant), and *population exposed* to sickness in people, according to the model of [6]. Given a budget of 5 sensors per network, we generated optimal solutions for each network and impact type.

Our computational results are found in Figures 3 and 4. Although the actual travel times of contaminated water do not match the predictions made by MORPH, these are sufficiently similar to obtain similar *impacts* for similar contamination events.



(a) (b)

Figure 2: The 97 node 'EPANET example 3' network drawn with coordinates obtained from multi-dimensional scaling with (a) the original pipe lengths, and (b) the morphed pipe lengths

Note the for the EPANET example, the optimal solution, given our modeling assumptions, is equivalent in the original and morphed networks. The sensor locations themselves were very close, as in this example small differences in node identifier indicate close proximity. The symmetry of the grid network makes it more likely that many sets of sensor locations may produce good results, and indeed, our sensor locations for the morphed model differ significantly with those for the original. However, the objective values realized by these two solutions are within 2% of one another, indicating that the spread of contamination is comparably responsive to 5 sensors in each network.

Network	Objective	Value	Events $\times$ Witnesses	Sensor Locations
Original	mass consumed	2566995	40400	56 61 76 4 16
Morphed	mass consumed	2539625	40400	83 65 68 76 41
Original	population exposed	53.84	40400	47 79 56 61 12
Morphed	population exposed	53.84	40400	9 100 5 12 25

Figure 3: Sensor placement results for the grid network. The "Value" (objective value) and "Events $\times$ Witnesses"columns indicate that contaminant spread in the morphed model follows roughly the same pattern as it does in the original. The latter column counts the number of nodes that experience nonzero concentration, summed over all events. Since there are symmetries in this network and only one flow pattern, many different sensor placements can achieve an optimal solution.

Network	Objective	Value	Events × Witnesses	Sensor Locations
Original	mass consumed	4001568153000	7731	2 4 59 71 81
Morphed	mass consumed	3725635521186	5201	2 14 59 70 81
Original	population exposed	329.68	7731	2 4 14 59 75
Morphed	population exposed	435.47	5201	2 4 15 59 81

Figure 4: Sensor placement results for the 97 node EPANET example 3 network. With the more complicated, varying flow patterns in this network, the morphed model does not match the contaminant spread properties of the original network as closely as with the grid network. However, these patterns are sufficiently similar to achieve nearly identical optimal sensor placements (in this network, similar node id's indicate nodes in close proximity to each other).

5 Conclusions

We have presented a nonlinear optimization model to morph water distribution network models by altering their pipe lengths and diamaters. For two small datasets, we have demonstrated that the original layouts cannot be determined from morphed networks using methods that do reconstruct the layouts, given the original network models. Furthermore, we found that the patterns of contaminant impacts during water quality simulations are sufficiently similar that sensor placements computed based on simulations on morphed networks are similar to those based on the original network.

It remains to be shown that this technique is scalable to large inputs. Better use of IPOPT may achieve this goal.

Along with sanitized node and pipe identifiers, this morphing technique is a first step toward securely sharable, yet realistic water network data for evaluating algorithms. We hope to spur a new sub-area of research within the water community to augment this work. Eventually, we hope to see realistic water network data more widely available to algorithm researchers.

Before we would recommend sharing data derived from real network models, further work should be done on network equivalence. Our MORPH model makes predictions for travel times that are not always realized, so modeling improvements may be necessary. Also, for security purposes, the network topology should be modified in such a way that is is structurally different from the original, yet retains similar impact patterns during water quality simulations.

6 Acknowledgements

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94Al85000.

References

[1] J. Berry, W.E. Hart, C.A. Phillips, J. Uber, and J.P. Watson. Sensor placement in municipal water networks with temporal integer programming models. *Journal of Water Resources Planning and Management: Special Issue on Drinking Water Distribution Systems Security (To Appear)*, 2006.

[2] A. Buja, D.F. Swayne, M. Littlman, N. Dean, and H. Hofmann. XGvis: Interactive data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics*, to appear.

[3] J. Eckstein, W. E. Hart, and C. A. Phillips. PICO: An object-oriented framework for parallel branch and bound. In D. Butnariu, Y. Censor, and S. Reich, editors, *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, pages 219–265. Elsevier Science Publishers, Amsterdam, The Netherlands, 2001.

[4] R. Fourer, D.M. Gay, and B.W. Kernighan. *AMPL: A Modeling Language for Mathematical Programming*. Duxbury Press, 2nd edition, 2002.

[5] W.E. Hart, J.W. Berry, R. Murray, C.A. Phillips, L. Riesen, and J.P. Watson. Spot: A sensor placement optimization toolkit for drinking water contaminant warning system design. In *Proceedings of the ASCE/EWRI Congress*, 2007.

[6] R. Murray, J. Uber, and R. Janke. Modeling acute health impacts resulting from ingestion of contaminated drinking water. *Journal of Water Resources Planning and Management: Special Issue on Drinking Water Distribution Systems Security (To Appear)*, 2006.

[7] L.A. Rossman. EPANET 2 user's manual. EPA/600/R-00/057, 2000.

[8] D.F. Swayne, D. Lang, and A. Buja. GGobi: Evolving from XGobi into an extensible framework for interactive data visualization. *Computational Statistics and Data Analysis*, 2002.

[9] A. Wächter and L.T. Biegler. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.