Multi-level automated sub-zoning of water distribution systems

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

Citation

As Published
http://scholarsarchive.byu.edu/iemssconference/2014/Stream-H/35/

Publisher
International Congress on Environmental Modelling & Software Society (IEMSS)

Version
Original manuscript

Citable link
http://hdl.handle.net/1721.1/111843

Terms of Use
Creative Commons Attribution-Noncommercial-Share Alike

Detailed Terms
http://creativecommons.org/licenses/by-nc-sa/4.0/
Multilevel sub-zoning of water distribution systems using graph partitioning approach

L. Sela Perelman*, M. Allen**, A. Preis***, M. Iqbal***, A. J. Whittle*

*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
(Email: linasela@mit.edu; ajwhittle@mit.edu)
**Faculty of Engineering and Computing, Coventry University, UK
(Email: michael.allen@coventry.ac.uk)
***Singapore-MIT Alliance for Research and Technology, Singapore
(Email: amipreis @smart.mit.edu; mudasser@smart.mit.edu)

Abstract
This paper presents a generic framework for improved analysis and management of water distribution systems that reduces the size of full-scale water distribution system (WDS) by partitioning the system into smaller sub-zones. The problem is to divide a WDS into balanced sub-zones (in terms of weights defined by the user, e.g., number of nodes, demand, and population) such that the number of inter-connecting edges is minimized. An unsupervised learning algorithm for graph partitioning adopted from graph theory is applied and tested for practical sub-zoning of WDS. Graph partitioning a flat partitioning algorithm for dividing a network with \( n \) nodes into \( k \) clusters, such that the total weight of edges crossing between clusters is minimized and the loads of all the clusters are balanced. The key contribution of the work is applicative dynamic, versatile, and computationally fast scheme for WDS sub-zoning. Results are demonstrated on a large water distribution system serving heavily populated areas in Singapore.

Keywords
Graph partitioning, Water distribution systems

INTRODUCTION
The water sector worldwide is facing growing challenges in providing adequate quantities and quality of water to a continuously growing population. The finite water resources, climate change, and population growth and shift to warmer climates exercise an increasing impact on water resources and the water industry. The creation of new water resources (e.g. desalinated water) involve high economic costs and environmental impacts, hence, conservation of water through efficient end-use and active loss-control have attracted much interest both in research and practice (Mutikanga et al., 2013). Water conservation, traditionally, tends to focus largely on the end user, e.g. installing water efficient fixtures in the home and the workplace (Kunkel, 2003). Whereas, water utilities traditionally operate without consistent standards for water accounting and water loss control.

Network sub-zoning is one of the tools for leakage and pressure management for water loss control. The requirement of sub-zoning is to define the properties of the sub-zones within a network (e.g. size limit, total demand), to identify their boundaries (i.e. pipes or valves), and to monitor these boundaries for leakage and/or pressure control (with a limited number of meters). For example, the management of district metered areas (DMAs) has proven highly successful for leakage management (Thornton et al., 2008; Kunkel, 2003). A DMA of a water distribution system is a specifically defined area, in which the quantities of water entering and leaving the district are metered (Morrison, 2004). The subsequent analysis of flow
calculates the level of leakage within the district. According to Kunkel (2003) up to 85% of the measured leakage in the UK has been eliminated through a national water loss control program based on DMA’s.

Urban water distribution systems (WDS) can reach a substantial size of hundreds to thousands of nodes (i.e. consumers) and links (i.e. pipes, valves). The layout of WDS is typically looped having multiple flow paths from the water sources to consumers. The looped layout of WDS, which provides a high level of reliability to the system supply in the event of mechanical failures (e.g. pipe breaks, valves malfunctions), imposes difficulties on water loss control. The partition of an existing water distribution system into sub-zones should consider size limits on each of the sub-zones, minimize the number of connecting pipes between sub-zones, and maintain the hydraulic performance and reliability of the system. The application of proposed sub-zoning methods to real large-scale water distribution systems has been found to be generally limited (Mutikanga et al., 2013). Finally, there is a lack of consensual quantitative measures for evaluating system partition, hence the results are generally analyzed qualitatively.

This work presents a generic framework for simplifying the full-scale WDS by partitioning the system into smaller, balanced sub-zones with minimum number of inter-connecting pipes/valves without the need to re-design the system. This study applies a network partitioning approach adopted from distributed computing. In graph theory, these algorithms aim at grouping similar or closely connected vertices such that the set of nodes in each group has better connections to the nodes belonging to the same group than to the remaining nodes in the network.

METHODS
The water distribution network can be naturally represented as a graph $G = G(V,E)$ over a set of vertices (nodes) $V$ and a set of connecting edges (links) $E$, where the vertices represent consumers, sources, and tanks and the edges – pipes, pumps, and valves. The graph is characterized by nodal weights $w_i, i \in V$ (e.g. demand, elevation), link weights $w_j, j \in E$ (e.g. diameter, flow), and an adjacency matrix $A$ based on network topology. The graph division to clusters is designated by the set $C = (c_1, \ldots, c_k)$. Many graph partitioning methods have been developed for analyzing complex networks (Schaeffer, 2007). In this work a graph partitioning approach is adopted to achieve balanced sub-zones with minimal number of inter-cluster edges in the WDS. The performance of the method is evaluated based on qualitative (i.e. visualization) and quantitative (e.g. number of inter-cluster edges) measures.

Graph partitioning
The problem of graph partitioning consists of dividing $n$ nodes of the graph into a predefined number $k$ of roughly equal sized clusters such that the number of edges connecting the clusters is minimal and typically it is desired that the cluster have equal size. Graph partitioning is a fundamental approach used in parallel computing, for allocating tasks to multiple processors so as to minimize the communications and equally distribute the computational burden among them. Many algorithms have been developed for the graph partitioning problem, mainly consisting of three classes of algorithms – spectral, geometric, and multi-level partitioning. In this work, a multi-level graph partitioning algorithm (Karypis and Kumar 1998) shown to provide better partitions than the spectral methods at lower computational cost for a variety of problems is used for the division of WDS to sub-zones.
The graph partition problem is solved by performing a sequence of bisections of the graph \( G = G(V, E); w_i \forall i \in V; w_j \forall j \in E \). Initially, a 2-way partition is obtained, and then each cluster is further partitioned using 2-way partition. Finally, after a series of partitions, a \( k \)-way partition of the graph is attained. To attain a computationally efficient bisection of the graph, the graph is reduced by aggregating its nodes and edges, the smaller graph is partitioned, and the original graph is then recovered to construct the final partition of the original graph. The main steps of the graph partition method are described below and can be found in more detail in Karypis and Kumar (1998).

**Graph partitioning algorithm main steps:**

Given a graph with weights on nodes and edges, i.e. \( G = G(V, E); w_i \forall i \in V; w_j \forall j \in E \), a multi-level partitioning involves:

1. **Coarsening** – the original graph \( G_0 \) is reduced into a sequence of smaller graphs \( G_1, \ldots, G_m \) by aggregating its nodes and edges, such that \( |V_0| > \ldots > |V_m| \). The nodes are grouped based on *heavy edge* matching. A matching \( M_i \) of a graph \( G_i \) is defined as a set of edges in which no two edges are incident to the same vertex. The matching \( M_i \) is detected by traversing the nodes of the graph and adding the highest weight link, incident to the node, to the matching set. This process is repeated until all nodes have been visited. The next coarser graph is constructed by aggregating the nodes connected by the edges in the matching set, i.e. \( G_{i+1} = G(V_{i+1}, E_{i+1}) \) is induced by \( M_i \). The weight of the new aggregated meta-node in the coarser graph is equal to the sum of weights of the grouped nodes and the new set of edges equals to the union of the edges connecting the grouped nodes.

2. **Partitioning** – the reduced graph \( G_m \) is partitioned into two equal size clusters \( C_m \). The partition is carried out by growing regions around starting nodes using breadth first search and constantly updating the weights of the regions and the weights of the inter connecting edges. The high weight edges are greedily selected during the search to reduce the total weight of the edges connecting the clusters in the final partition.

3. **Recovering and refining** – the original graph \( G_0 \) is recovered from the partition \( C_m \). The recovery of the original graph can be attained simply by going through the intermediate partitions \( C_m, C_{m-1}, \ldots, C_0 \) and at each level \( i \) recover the partition of the original graph \( G_m, G_{m-1}, \ldots, G_0 \) by assigning the set of nodes grouped in each coarsening level \( i \). During each recovery level, a local refinement heuristics is used to improve the partition \( C_i \) of the un-grouped graph \( G_i \). This is achieved by iteratively swapping nodes between two clusters that reduce the weight of the cut edges.

The graph partitioning algorithm results in a single partition of the WDS with balanced sub-zones connected by a minimal number of links between the sub-zones. The implementation of the partitioning algorithm to WDS requires the definition of network graph, weights for nodes and links of the graph, and the number of desired sub-zones. The number of sub-zones can be inferred from the desired size of the sub-zones.

**Quality measures**

Evaluating the quality of a clustering algorithm or comparing different clustering methods is a difficult task. Mainly because the correct clustering is unknown, clustering algorithms rely
on different data sets, and their performance is dependent on parameter selection. Several qualitative and quantitative measures exist to evaluate the quality of the clustering (Schaeffer, 2007). The measures for evaluating the clustering to achieve balanced sub-zones with minimal number of inter-cluster edges (where flow is to be measured) for water loss management are:

1. Adjacency matrix – visualization of the adjacency matrix is a graphical measure for evaluating the quality of the clustering. When the nodes of a graph are ordered randomly, there is no apparent structure in the adjacency matrix. Re-ordering of the nodes according to their clusters should reveal a tight block-diagonal structure of the adjacency matrix.
2. Cluster diagram – the layout of the clustering of the original graph can qualitatively assist in the evaluation of the clustering of the WDS.
3. Total cut-size – the total number of inter-cluster links connecting the different clusters implies the number of links that need to be monitored for water loss control i.e. this constraint defines the number (and cost) of sensors that need to be installed across the network. Naturally, this number grows with the number of desired sub-zones and should be minimized.
4. Worst cut-size – this measure amounts the total number of links that need to be monitored for a specific cluster. This number should also be minimized to limit the dependencies between the different zones of the WDS.
5. Cluster size – for better control of the WDS it is desired that the load of each sub-zone will be roughly equal. The load is ultimately specified by the water utility and can be measured in terms of the estimated demand, population served, and/or number of connections.

APPLICATION
The suggested approach for sub-zoning was tested on a network serving part of Singapore. The graph partitioning algorithm was implemented in METIS (METIS, 2013) using the gpmetis function. The adjacency matrix of the network and weights of its links and nodes are a required input for the graph partitioning algorithm. The weights of the links were uniform and nodal daily demands were assigned to network nodes. The network was partitioned according to six demand loading constraints for each sub-zone, i.e. 20, 10, 8, 4, 3, 2% of the total daily demand of the network. The number of clusters was computed according to the demand load constraints, i.e. \( k = 5, 10, 15, 25, 35, 50 \), with \( k \) being the number of sub-zones. The required input is network topology, weights of nodes and links. The features of the program allow to select the desired clustering algorithm and to set size constraints or number of clusters. The results of the clustering of the network are presented in tabular and graphical schemes.

Table 1 demonstrates the output given by the approach for a partition to 25 sub-zones. For each sub-zone, the number of nodes, of intra and inter cluster edges, and the daily demand are listed in Table 1.

Next we compare and evaluate quality measures of the performance of the graph partitioning algorithm. Figure 1 shows the adjacency matrices of the – (a) original network and (b) – (d) network divided into 5, 10, and 25 sub-zones, respectively. The columns and rows of the matrix are reordered corresponding to the sub-zones represented by the blocks of the matrix. A clear cluster structure of the network can be observed from Figure 1(b)-(d) compared to the original structure Figure 1(a). It is important to note, that the matrix blocks represent the size of the clusters in terms of number of nodes and do not represent the demand of each cluster,
since the adjacency matrix represents the connectivity between the nodes and does not depict the weights of the nodes.

Table 1. Graph partitioning results for 25 sub-zones.

<table>
<thead>
<tr>
<th>#Sub-zone</th>
<th>#In-nodes</th>
<th>#Intra-edges</th>
<th>#Inter-edges</th>
<th>Daily demand [m³/day]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>124</td>
<td>125</td>
<td>2</td>
<td>653.88</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>45</td>
<td>4</td>
<td>567.97</td>
</tr>
<tr>
<td>3</td>
<td>74</td>
<td>76</td>
<td>5</td>
<td>545.28</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>50</td>
<td>3</td>
<td>662.91</td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>27</td>
<td>3</td>
<td>620.63</td>
</tr>
<tr>
<td>6</td>
<td>56</td>
<td>55</td>
<td>5</td>
<td>581.3</td>
</tr>
<tr>
<td>7</td>
<td>34</td>
<td>33</td>
<td>3</td>
<td>554.12</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>18</td>
<td>1</td>
<td>614.56</td>
</tr>
<tr>
<td>9</td>
<td>19</td>
<td>18</td>
<td>3</td>
<td>631.04</td>
</tr>
<tr>
<td>10</td>
<td>125</td>
<td>124</td>
<td>2</td>
<td>783.74</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>527.7</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>511.72</td>
</tr>
<tr>
<td>13</td>
<td>107</td>
<td>107</td>
<td>4</td>
<td>851.05</td>
</tr>
<tr>
<td>14</td>
<td>212</td>
<td>219</td>
<td>5</td>
<td>678.16</td>
</tr>
<tr>
<td>15</td>
<td>117</td>
<td>119</td>
<td>4</td>
<td>864.19</td>
</tr>
<tr>
<td>16</td>
<td>215</td>
<td>222</td>
<td>7</td>
<td>790.3</td>
</tr>
<tr>
<td>17</td>
<td>43</td>
<td>45</td>
<td>3</td>
<td>642.06</td>
</tr>
<tr>
<td>18</td>
<td>95</td>
<td>99</td>
<td>3</td>
<td>730.6</td>
</tr>
<tr>
<td>19</td>
<td>98</td>
<td>102</td>
<td>4</td>
<td>563.78</td>
</tr>
<tr>
<td>20</td>
<td>223</td>
<td>225</td>
<td>3</td>
<td>1110.54</td>
</tr>
<tr>
<td>21</td>
<td>111</td>
<td>114</td>
<td>3</td>
<td>570.36</td>
</tr>
<tr>
<td>22</td>
<td>100</td>
<td>99</td>
<td>2</td>
<td>603.62</td>
</tr>
<tr>
<td>23</td>
<td>244</td>
<td>248</td>
<td>5</td>
<td>784.97</td>
</tr>
<tr>
<td>24</td>
<td>97</td>
<td>98</td>
<td>6</td>
<td>724.96</td>
</tr>
<tr>
<td>25</td>
<td>193</td>
<td>197</td>
<td>7</td>
<td>814.76</td>
</tr>
</tbody>
</table>
Figure 1. Connectivity matrix: (a) original network and (b) – (d) network divided into 5, 10, and 25 sub-zones, respectively.

Figure 2 demonstrates the performance of the partitioning algorithm for six different divisions based on three suggested quantitative measures: (a) total cut-size, (b) worst case cut-size, and (c) cluster size. From the results it can be seen, that, as expected, the total number of inter-cluster connecting links grows with the number of sub-zones, the demand load of each sub-zone decreases, and the maximum number of inter-cluster connecting links for a single sub-zone varies around 9.

Additional outcome from the partition of WDS to sub-zones is the cluster structure of the network. Figure 3 (a) demonstrates the structure the network after partition to 20 sub-zones and the connections between the different zones and the network sources. The number on the edges shows the number of inter-cluster connecting links and the direction is shown for a representative daily flow pattern of the distribution system. Figure 3 (b) graphically shows map view of the network.

Finally, Figure 4 shows an example of how the sub-zoning could be utilized for better analysis and management of the WDS. Figures 4 (a)-(b) demonstrate the average demands and pressures of each sub-zone. It can be seen that sub-zone 29 has relatively low pressures during the day. From analyzing the network, it was seen that the sub-zone is connected by two 150 [mm] pipes, which results in high head-losses.
Figure 2. Quality measures as a function of the number of sub-zones (i.e. 5, 10, 15, 25, 35, 50): (a) total number of inter-connecting edges, (b) worst case of number of inter-connecting edges for a single sub-zone, and (c) sub-zone demand

Figure 3. Partitioning to 20 sub-zones: (a) Block diagram and (b) Map view
CONCLUSIONS
The suggested framework is used to simplify a WDS structure by organizing the water consumers in (virtual) sub-zones. The framework is flexible and easily adoptable to the needs of the water utility such as selecting different objective measure (e.g. population of a zone) and setting different weights to network nodes and links to adequately represent the objective. The performance of the algorithms was evaluated and compared using qualitative and quantitative measures. It was shown that the methods are compatible and applicable for large-scale WDS. The suggested methods can provide a decision support tool to water utilities for network sectorization for better management of water supply.

REFERENCES