Hybrid Evolutionary Optimization/Heuristic Technique for Water System Expansion & Operation

Mark Morley[*1] & Carla Tricarico[2]

[1]*Centre for Water Systems, University of Exeter, Exeter, Devon, EX4 4QF (United Kingdom) – Research Fellow*
[2]*Dipartimento di Ingegneria Civile e Meccanica, Università degli Studi di Cassino e del Lazio Meridionale, via Di Biasio 43, 03043 Cassino (Italy) – Assistant Professor*

Abstract

This paper presents a methodological solution to The Battle of Background Leakage Assessment for Water Networks (BBLAWN) competition. The methodology employs two constrained multiple-objective optimization problems and is implemented in the context of a software application for the generic hydraulic optimization and benchmarking of Water Distribution System (WDS) problems. The objectives are the combined infrastructure and operational costs and system-wide leakage, both to be minimized. In order to accelerate the evaluation of potential solutions, a distributed computing approach permits multiple EPANET solutions to be evaluated in parallel. A pressure-driven demand extension to EPANET assists the optimization in accurately ranking near-feasible solutions and to dynamically allocate leakage demand to nodes. Pressure Reducing Valves (PRVs) have been located in two ways: *a priori*, with respect to the optimization analysis and *a posteriori* after the infrastructure optimization to reduce excess pressure and pipe leakage. The latter demonstrates better overall fitness, leading to optimal configurations dominating those obtained with the former. Several temporal resolutions for PRV settings have been evaluated to contrast the optimal solutions with the computational effort required.

Author Keywords: Multiple Objective Optimization; Evolution Algorithms; BBLAWN; Water Distribution Systems; Leakage

Introduction

The Battle of Background Leakage Assessment for Water Networks competition (BBLAWN) (Giustolisi *et al.*, 2015) presents a challenge in optimizing a water network both in terms of design/expansion and also operation. The analysis requires the reinforcement through the replacement or augmentation of its components (tanks, pipes, etc.) and better management of the system operation by regulating the use of pumps and the installation of PRVs in order to minimize leakage and operational costs in terms of energy consumption.

Population-based optimization techniques have emerged over the last few decades as a popular technique for application to water distribution system design, operation and rehabilitation problems. Within this class of technique, a number of approaches have been proposed including genetic algorithms (Savić and Walters, 1997) and memetic algorithms such as the Shuffled Frog Leaping Algorithm (Eusuff and Lansey, 2003) and Ant Colony Optimization (Maier *et al.*, 2003). A number of these have been applied to the BBLAWN problem in order to determine which might be the most effective approach (Morley and Tricarico, 2014). The high dimensionality of the problem appeared to cause many of the techniques to struggle with the optimization – leading to the selection of the Omni-Optimizer algorithm (Deb and Tiwari, 2008) coupled with both in-process and post-processing heuristics – which produced the runner-up solution in the competition. In this paper, the analysis has been refined by analysing in greater detail the optimal component of the system and by allocating the location of PRVs in the system by means of two different approaches, *a priori* with respect to the optimization runs or *a posteriori* in which an initial optimization of the infrastructure with no PRVs installed is undertaken. This latter method has demonstrated solutions characterized by a better fitness with respect to those obtained by the original methodology applied by Morley and Tricarico (2014).

Methodology

The software methodology employed combines a revised, pressure-driven version of EPANET (Morley and Tricarico, 2008; Rossman, 2000) with a C++ implementation of optimization algorithm to model the effect on the hydraulic network performance under the varying system parameters derived through the optimization process. This combination is embodied in a unified, generic WDS optimization application, also developed in C++.

Objectives

The BBLAWN optimization has been formulated as twin-objective optimization problem to minimize:

1. Total Cost – the sum of annualized infrastructure upgrade costs (pipe replacement and duplications, tank, pump and valve installation) and annual operational (pumping) costs.
2. Leakage – the absolute annual volume of water lost as leakage.

Hydraulic Solver

The BBLAWN problem introduces a leakage model whereby leaks are calculated on a per-pipe basis and then aggregated into the demand nodes as per Giustolisi, *et al.* (2015).

Since the leakage ascribed to a particular node is a function of the pressure both at itself and at the nodes at the end of each attached link, it is not possible to use the standard EPANET emitter component to model the leakage which operates on the basis of the available pressure at a single node. One approach would be to run the EPANET model normally and then adjust the demands to account for the leakage and to rerun the model repeatedly until convergence was reached. Whilst this has the advantage of not requiring any modifications to EPANET directly, it was discounted because of the extended run-times that such a strategy would necessarily entail.

Having successfully retrofitted a pressure-driven extension to EPANET previously (Morley and Tricarico, 2008) the authors have experience in adapting and extending the hydraulic solver and, accordingly, the leakage model described above has been incorporated directly into the C language source code of the EPANET toolkit. A number of functions have been modified (detailed in Table 1) to accommodate the leakage model as part of the normal iterative cycle employed by EPANET to produce the hydraulic solution. In addition, further variables were added to EPANET in order to store the leakage parameters alpha and beta for each link as well as the calculated leakage on a per-link and per-node basis. This approach has the advantage that by directly manipulating the solution matrices employed by EPANET, it is relatively straightforward to allocate leakage to tanks (as is required according to the rules). Ordinarily, EPANET does not allow the direct assignation of demands to tanks as would be necessary in this instance – requiring the introduction of additional dummy nodes and pipes in order to model this leakage correctly.

TABLE 1 TO BE INSERTED HERE

The use of EPANET with a stochastic optimization process commonly results in a large number of hydraulically-infeasible solutions being generated and subsequently evaluated by the hydraulic solver. The evaluation of these infeasible solutions takes additional time as, typically, the maximum number of solver iterations is expended attempting to converge the model and, additionally, large numbers of intermediate timesteps may be introduced into the evaluation. The algorithm used seeks to avoid the worst impacts of infeasible solutions by terminating their execution after the first timestep in which they demonstrate hydraulic infeasibility. Instead of penalizing the solution heavily in order to hasten its departure from the population, the solution is marked as infeasible and estimates of its constraint violations are extrapolated, weighted by the proportion of the extended period simulation that had been successfully completed prior to the infeasibility. This results in a commensurate reduction in the runtime "wasted" in evaluating infeasible solutions as well as preserving the genetic diversity of the population to the maximum extent possible.

Optimization Environment

The software presented in this paper also includes a distributed-processing system in order to militate against the extended runtimes that are a common issue when optimizing with evolution algorithms. The BBLAWN optimization is characterized by a particularly high number of decision variables as seen in Table 2. As a consequence of this, the deEPANET system (Morley et al., 2006), which employs the industry standard Message Passing Interface (MPI) protocol to parallelize the hydraulic simulation computation, was incorporated into the methodology. This system permits the concurrent evaluation of a large number of potential solutions either on local processors or to other computers on a LAN. Owing to the relatively long runtimes of the hydraulic simulations compared to the data transfer speeds across a modern Gigabit LAN, near linear improvements in GA runtime are achievable as processing cores are added to the cluster. For the purposes of this optimization the software was deployed across a cluster of three workstations, each equipped with two Intel Xeon E5645 CPUs packages which comprise six cores running at 2.4 GHz.

Decision Variables

For the purposes of optimizing the BBLAWN problem, no attempt was made to simplify the problem. Legitimate approaches to doing this might have included grouping adjacent pipes with similar characteristics or restricting the application of the optimization to pipes over a given length. Table 2 enumerates the configuration of the decision variables used in the optimization. In the first instance, as in Morley and Tricarico (2014), the potential sites for the 39 possible PRV installations were determined through engineering judgment prior to starting the optimization and, naturally, this will have biased the range of potential solutions, accordingly. The resolution afforded the settings of the PRVs has been considered (Table 3) and separate optimizations have been undertaken for four different schemes: one single fixed setting for each PRV for the entire simulation; a daily variation in which it has been assumed a different setting for each day of the simulation– i.e. 7 values for each PRV in total; one value for every 6 hours of the simulation– i.e. 28 values for each PRV; and one value for each hour of the simulation the maximum resolution permissible under the rules of the competition giving 168 settings for each PRV.

As a second stage, as reported below, the problem has been reformulated as a two-stage optimization in which PRV locations have again been placed according to engineering judgment but following an initial optimization without considering PRVs which determines the optimal infrastructure arrangement *a priori*. With the network so optimized, the valves have been located using two criteria: (1) available head for pressure reduction (i.e. making zones out of nodes that have significant excess pressure for the majority of the simulation); (2) Analysis of the maximum quantity of downstream leakage that can be reduced in order to save money. This is achieved by assuming that the each node in the network can be reduced to a theoretical minimum pressure of 20m – the minimum permissible. This allows the quantification of a maximum amount of leakage that can be saved for each pipe. With this further analysis the number of PRVs to be located in the network have been reduced to 33. The pressure-setting for each PRV has been considered as detailed previously (i.e. fixed, daily, every 6h and every 1h).

TABLE 2 TO BE INSERTED HERE

TABLE 3 TO BE INSERTED HERE

Constraints

During the evaluation of potential solutions a number of "hard" constraints are employed to ensure that the solution under consideration meets the minimum criteria to be considered as a solution. The constraints are divided into those general constraints which are applicable to all such optimizations such as hydraulic feasibility and avoiding negative pressures, disconnected nodes and pumps operating outside their normal flow regime. In addition there are a number of problem-specific constraints for the BBLAWN optimization, comprising: all demand nodes with a demand meeting a minimum pressure requirement of 20m; tanks not being permitted to empty at any time through the simulated time horizon and the final levels of tanks being at least as high as their initial levels to ensure that a solution is repeatable over successive weeks. Differential constraint weightings are used to signify the relative importance of meeting the optimization constraints. The

EPANET Error and EPANET Warning constraints are given the highest priority in order to prioritise the generation of feasible solutions by the optimization. Solutions which violate hard constraints are considered unfeasible by the optimization algorithms and as such are unlikely to play a significant role in the evolution of the population once more favourable, feasible solutions have been identified.

Inline heuristics

The formulation of the BBLAWN problem includes a pricing differential between the cost of replacing the pipe and the installation of a duplicate, parallel pipe. This is realised as a premium of 20% or the parallel pipe, ostensibly to cover the additional costs of installing an entirely new pipe. As detailed above, the optimization has complete freedom to select either replacement (including closure) and/or duplication options for each existing pipe. Accordingly, a number of heuristics were added to the objective function to ensure that the most cost-effective option is selected in each instance. These heuristics include:

- If a pipe is to be closed and also duplicated, then the selected duplicate pipe diameter is chosen as a replacement pipe – given that this will necessarily be 20% cheaper to install.
- If a pipe is to be duplicated as well as replaced, and the selected duplicate pipe diameter is larger than the replacement (and is therefore more expensive), the pipe diameters are reversed so that it is the cheaper pipe that attracts the 20% premium.
- If a pipe is to be duplicated and the existing pipe is not to be closed, a test is made to see if it is more cost-effective to install a single pipe with the same or greater cross-sectional area to the two pipes combined.

Post-processing heuristics

Owing to the very high dimensionality of the problem as formulated, it was considered likely that there would be scope to further improve the quality of the solutions obtained during the evolutionary algorithm phase of the optimization. To that end two heuristics are applied to the resulting solutions in order to identify feasible, incremental improvements can be applied to a given solution. The two heuristics operate in a mutually exclusive fashion and can be repeated a number of times.

In order to reduce the installation cost of the pipe infrastructure, at the expense of available pressure, the first heuristic attempts to reduce, sequentially, the pipe diameters in the network. The heuristic operates recursively from the extremities of the network, inward, and can be seen to work well for purely dendritic networks. In the event of the recursion encountering a loop, each branch of the loop is evaluated separately in turn and the most cost-effective combination implemented.

The second heuristic varies (downward) the pressure settings of each of the PRVs in the network for each timestep in the simulation in an attempt to further reduce available pressure and thus reduce the pressure-dependent leakage accordingly. PRVs are considered for this reduction in the order of the highest differential from upstream to downstream, for each timestep.

Discussion of Results

Issues

In contrast to the previous Battle problem (Marchi et al., 2014; Morley et al., 2012), the outputs of the hydraulic solver do not need to be directly compared with a reference version of the EPANET solver. As a consequence, minor variations in the computation are no longer as critical for assessing the suitability of the proposed solution. However, the scale of the unconstrained problem as described above has introduced further challenges related to memory capacity. 32-bit computers are limited to accessing 4GB of memory whilst 32-bit operating systems may introduce further constraints – in the case of Microsoft Windows, each process may access a maximum of around 1.6GB. The unconstrained problem, as outlined above, requires a greater amount of memory, particularly when considering the large population sizes that the algorithms under consideration require when contemplating such large decision spaces. In order to consider a full evaluation of the unconstrained problem, using all of the decision variables, it proved necessary to move to a 64-bit implementation of the software to avoid this process limit. As has been seen previously with variations between single and double-precision versions of EPANET, the move to a 64-bit version revealed appreciable differences between the numeric solutions achieved with the 32-bit version. It is thought that these numerically minor variations are present as a consequence of differing standard libraries being employed by the 32-bit and 64-bit varieties of C++ being employed. For the purposes of the analysis herein, all results were evaluated using a 64-bit, double precision version of the EPANET solver.

Accurately establishing pump energy consumption is somewhat problematic using the EPANET toolkit API. Instead of returning an average consumption (or total consumption) over the reporting timestep, the EN_ENERGY result returns an instantaneous value for energy consumption. As a consequence of this, retrieving the total energy consumption for a network which has many state changes (introducing intermediate timesteps) is a somewhat contrived process. Therefore, it is necessary to recalculate both energy consumption and leakage at all intermediate timesteps in a simulation in order to obtain accurate values for both.

Result overview and comparison

For the most part, it can be seen that the optimal solutions preferred for the BBLAWN problem using this methodology are characterized by the replacement of most or all of the pipes in the network and a very small number of, or no, pipe duplications. This is a sensible outcome given the 20% cost penalty associated with pipe duplications. A more surprising and common feature in the results is the absence of any supplementary tank storage. All of the optimization techniques tested in this study employed enlarged tanks in the early stages of their evolution but later were seen to remove these from later solutions as the optimizations progressed.

A priori PRV optimization

By considering the *a priori* allocation of PRVs in the system for 39 PRVs, the Pareto Fronts obtained by varying the PRVs setting are illustrated in Figure 1. Table 4 reports the summary of the results obtained for each optimization, considering the solution with minimum total cost and minimum leakages (the solutions circled on each pareto front in Figure 1). Given the absence of any reliability criterion, it is not surprising that the GA opted for dendritic network forms, removing all but one loop from each of the candidate solutions in the resultant populations. It is interesting also to note that, in contrast to the results obtained in Morley and Tricarico (2014) these optimizations have not preferred not to isolate the tank T6.

FIGURE 1 TO BE INSERTED HERE

TABLE 4 TO BE INSERTED HERE

From the results reported the lowest total cost solution is that of the single fixed value for each PRV. The results for the optimizations with greater degrees of freedom for the PRV settings compare unfavorably. In particular, the optimization with hourly settings for the PRVs struggled to match the objective values of the other runs – likely owing to the significantly larger search space associated with this configuration.

A posteriori PRV optimization

Along with the single stage optimizations, outlined above, a two stage optimization was also formulated in which location of the PRVs is determined through expert judgment following an initial optimization to determine the optimal infrastructure. A solution has been selected from the resulting Pareto front, representing the lowest overall cost solution (€1,314,874). A number of thematic maps (for example, Figure 2) were generated to assist in the placement of the PRVs, including quantification of total leakage from each pipe and the mean surplus pressure experienced at each node. Prior to PRV installation, this network configuration experiences annual leakage, by value, of €609,520. By considering a theoretical scenario in which the pressure can be reduced to the minimum required of 20m at each node, it is possible to place a lower bound on the leakage for this particular network configuration, amounting to €240,823/year. Expert analysis yielded 33 pressure zones for installation – reduced from the 39 potential pressure zones identified at the outset and employed in the other optimization strategy. This reduction can be attributed to the simplification of the network into a dendritic form and a better understanding of the distribution of surplus pressure within the network. The dendritic form, whilst cheaper to implement, may be considered a less reliable topology – an objective not considered by the BBLAWN optimization.

FIG. 2 TO BE INSERTED HERE

The Pareto Fronts resulting from the *a posteriori* analysis undertaken (Figure 1) demonstrate lower total costs than those obtained previously with the *a priori* PRV allocation, employing at the same time a reduced number of decision variables, reducing the computational effort required. The Summary of the optimal results obtained have been reported in Table 5 in which at the "no PRVs" solution have been compared the optimal solutions obtained by letting fixed or vary the pressure settings of the PRV as before.

TABLE 5 TO BE INSERTED HERE

Conclusions

A novel methodology for the expansion and operation optimization problem for the BBLAWN case study has been applied and solved by means of a population-based algorithm incorporating heuristics both within the optimization process and in post processing. The problem has been considered has a two objectives one in which it was necessary to minimize both operational/design costs and leakages. The BBLAWN leakage model has been directly incorporated into a pressure driven extension of EPANET hydraulic solver to maximize the efficiency of the leakage evaluation. Evaluation of the problem has been distributed on a local cluster computing resource using the deEPANET software for parallelizing the hydraulic simulations associated with each individual solution generated by the optimization. The analysis of the problem in order to respect the "battle" criteria has been solved by means of a methodology based on engineering judgment supported by the optimization algorithm. The problem has been solved by means of two different approaches in which the PRVs location have been located *a priori* respect to the optimization or *a posteriori* following an initial optimization of the infrastructure alone. The latter analysis has demonstrated better solutions for both of the objectives under consideration. In addition, different pressure-setting schemes for the PRVs have been considered although the results demonstrate that there is an insignificant difference in terms of objective values achieved if a fixed pressure setting is assumed compared to those schemes which required greater computational effort.

As a recommendation for future work, a better cost model would consider the cost and reliability implications of pump status/valve setting switches and would allow the optimization to attempt to minimize this type of cost.

References

Deb, K. and Tiwari, S. (2008). "Omni-Optimizer: A Generic Evolutionary Algorithm for Single and Multi-Objective Optimization", *European Journal of Operational Research*, **185**(3), pp.1062-1087.

Eusuff, M.M. and Lansey, K.E. (2003). "Optimization of Water Distribution Network Design Using the Shuffled Frog Leaping Algorithm", *Journal of Water Resources Planning and Management* - ASCE, **129**(3), pp.210-225.

Giustolisi, O., Berardi, L., Laucelli, D., Savić, D., Kapelan, Z. (2015). "Operational and Tactical Management of Water and Energy Resources in Pressurized Systems: the competition at WDSA2014" *Journal of Water Resources Planning and Management*, Special Issue on The Battle of Background Leakage Assessment for Water Networks [doi: 10.1061/(ASCE)WR.1943-5452.0000583]

Maier, H.R., Simpson, A.R., Zecchin, A.C., Foong, W.K., Phang, K.Y., Seah, H.S., Chan Lim Tan, H.S. (2003). "Ant Colony Optimization for Design of Water Distribution Systems", *Journal of Water Resources Planning and Management* - ASCE, **129**(3), pp.200-209.

Marchi A., Salomons E., Ostfeld A., Kapelan Z., Simpson A., Zecchin A., Maier H., Wu Z., Elsayed S., Song Y., Walski T., Stokes C., Wu W., Dandy G., Alvisi S., Creaco E., Franchini M., Saldarriaga J., Páez D., Hernández D., Bohórquez J., Bent R., Coffrin C., Judi D., McPherson T., van Hentenryck P., Matos J., Monteiro A., Matias N., Yoo D., Lee H., Kim J., IglesiasRey P., MartínezSolano F., MoraMeliá D., RibellesAguilar J., Guidolin M., Fu G., Reed P., Wang Q., Liu H., McClymont K., Johns M., Keedwell E., Kandiah V., Jasper M., Drake K., Shafiee E., Barandouzi M., Berglund A., Brill D., Mahinthakumar G., Ranjithan R., Zechman E., Morley M., Tricarico C., de Marinis G., Tolson B., Khedr A., Asadzadeh M. (2014). "Battle of the Water Networks II", *Journal of Water Resources Planning and Management* - ASCE, **140(7)**, pp. 1-14.

Morley, M.S., Tricarico, C., Kapelan, Z., Savić, D.A. and de Marinis, G. (2006). "deEPANET: A Distributed Hydraulic Solver Architecture for Accelerating Optimization Applications Working with Conditions of Uncertainty". In: GOURBESVILLE, P., CUNGE, J., GUINOT, V. & LIONG, S-Y. (eds.) *Proceedings 7th International Conference on Hydroinformatics*, Nice, France, pp.2465-2472.

Morley, M.S. and Tricarico, C. (2008). *Pressure Driven Demand Extension for EPANET (EPANETpdd)*, Techincal Report 2008-02, Centre for Water Systems, University of Exeter, UK, 11pp. DOI:10.13140/2.1.2094.9760

Morley, M.S., Tricarico, C. and de Marinis, G. (2012). "Multiple Objective Evolutionary Algorithm Approach to Water Distribution System Model Design", *Proceedings WDSA 2012 – 14th Water Distribution Systems Analysis Conference*, Adelaide, Australia, pp.551-559.

Morley, M.S. and Tricarico, C. (2014). "A comparison of population-based optimization techniques for water distribution system expansion and operation", *Procedia Engineering*, **89**, pp13-20. DOI:10.1016/j.proeng.2014.11.154.

Rossman, L.A., *EPANET 2 User's Manual*, United States Environmental Protection Agency, Washington, USA, 2000, 200pp.

Savić, D.A. and Walters, G.A. (1997). "Genetic algorithms for least-cost design of water distribution networks", *Journal of Water Resources Planning and Management* – ASCE, **123**(2), pp.67-77.

List of Tables:

List of Figures: