WDSA CCWI 2022

2nd International Joint Conference on Water Distribution Systems Analysis & Computing and Control in the Water Industry
Valencia (Spain), 18-22 July 2022
doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14136

DATA DRIVEN APPROACH FOR EQUITABLE SUPPLY IN WATER NETWORKS

Adhityan R[1], Varghese Kurian[2] and Sridharakumar Narasimhan[3]

[1,2]Indian Institute of Technology Madras, Chennai, Tamil Nadu, (India),
[3]Dept. of Chemical Engineering and Robert Bosch Centre for Data Science and Artificial Intelligence, Indian Institute of Technology Madras, Chennai, Tamil Nadu, (India)

[1]*nsraadhi@gmail.com*, [2]*varghese@alumni.iitm.ac.in*,
[3]*sridharkrn@iitm.ac.in*

Abstract

In many localities with limitations on the availability of water, Water Distribution Networks (WDNs) supply water only for a few hours in a day. In such networks, inefficient operational policies can lead to inequitable supply of water. This work proposes a data-driven approach for scheduling the supply side of WDNs for equitable distribution of available water. We formulate and solve a scheduling problem that makes use of flow measurements from the real network rather than a hydraulic model of the WDN. This helps reduce the effort required for model development and the errors arising out of it. Further, to limit the number of additional measurements required, we propose a heuristic for choosing measurements that are informative. In each step of this iterative procedure, a schedule is generated using the available measurements and based on this, a new system state that has to be measured is identified. The iteration is completed when acceptable performance is achieved. We demonstrate the advantages of this technique through simulations of a real WDN and experiments performed on a lab-scale network.

Keywords

Water distribution networks (WDN), Equitable distribution, Scheduling algorithms, LabVIEW.

1 INTRODUCTION

Water Distribution Networks (WDNs) are used for transporting water from source to consumers. These systems are composed of reservoirs, pipes, pumps, and other instrumentation. The flow rates in these systems may not be proportional to the demands at the consumers' side. In such situations, utility providers control the valves and pumps in the system to distribute water efficiently and equitably. However, the mathematical program for scheduling valves and pumps in a WDN is a difficult problem to solve (1). Recently, in (2), the authors proposed an efficient technique for scheduling a class of rural WDNs. However, the technique requires a substantial amount of data from the system to develop the schedule. In the present work, we propose a technique for judiciously choosing experiments, leading to the generation of useful data, which in turn, can be used for scheduling the system efficiently.

Scheduling WDNs is a complex problem to solve and for the same reason, a variety of formulations and algorithms have been proposed for solving these efficiently (3; 4; 5). However, only a handful of these techniques address the drinking water networks widely prevalent in developing countries. Except for a few recent works, the majority of the research addresses urban water networks (6; 7; 1). Among these, the scheduling technique proposed by Amrutur et al. (6) is computationally expensive, the method of Bonvin et al. (1) applies only to systems with continuous control valves and the improved algorithm presented by Bonvin et al. (8) (also (1; 6)) requires the full specification of the WDN in consideration. Moreover, none of these techniques address the problem of equitable distribution, a key challenge in rural drinking water networks (9). While the technique proposed by Kurian et al. (2); (henceforth referred to as $\mathcal{A}_1$) presents an efficient technique for equitable distribution of water, the technique is data intensive. That is, the number of measurements required for optimal scheduling increases exponentially with the number of valves in the system. Therefore, it is desirable to have a method for optimal experiment design that would generate informative data for the scheduling problem.

The literature addressing the estimation problems in water networks primarily focus on identifying the Hazen William's coefficient, system demands or the pump characteristics for model calibration (10; 11). These works identify the model parameters that are to be adjusted, choose the data for model calibration and finally present methods for identifying the parameters. Here, the objective is to identify a system model that is predictive of all characteristics of the system (pressures, flow rates at all locations) arising at any operating condition of the system. However, in the context of scheduling problems in drinking water networks, what we are really interested in are the *flow rates received by the demand locations under the control settings that the system is expected to be operated in*. A tailor-made design of experiments could potentially help us achieve the objective with lesser data. This is the key idea distinguishing our work from existing literature on identification in WDNs.

The rest of the paper is organised as follows. We present a brief review of the scheduling technique $\mathcal{A}_1$ and highlight the problem of optimal experiment design that is addressed in the present work. Following this, we present a method for solving the problem along with a brief analysis of the algorithm. The implementation is demonstrated on simulated networks and a laboratory scale test network. We conclude with the key takeaways and recommendations for future work.

2 PROBLEM STATEMENT

We address the problem of equitable distribution in a class of rural drinking water networks. *Figure 1* shows the schematic of one such network, originally presented by Bhave and Gupta (12). In these systems, water is supplied from a single source to a set of downstream consumer locations (village level storage tanks/standpipes) using pumps or gravity. As there are storage facilities available at the consumer locations, water may be provided at any time of the day to meet the daily demand. Further, it is assumed that there is a valve upstream of every consumer location which can be turned ON or OFF to regulate the supply.

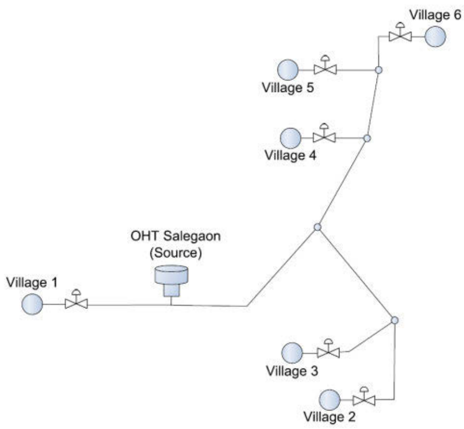

Figure 1. Schematic of the Osmanabad WDN (12)

The objective of the scheduling task is to identify the time points for operating the valves to distribute the available water equitably. For this, the objective function in the mathematical program minimises the deviation of the supplied water from the daily demand of the respective consumer location.

In algorithm $\mathcal{A}_1$, the authors decoupled the system model from the optimisation problem. As the state space for the system is finite (due to the finite number of valve settings), the flow rates in the system under different valve configurations were measured *a priori*, and these measurements were used for solving the optimisation problem for identifying the schedule (*Figure 2*). This decoupling allows the transformation of the scheduling problem - which has to be solved on a daily basis - from a non-linear program (P) to a linear problem ($P1$), at the expense of a set of experiments (or model simulations) that have to be performed only on the first day. As the only measurements required are the flow rates at the final consumer locations, the technique could be implemented even without any prior knowledge of the network topology. However, the number of experiments required to explore the entire state space rises exponentially with the number of valves and this could become prohibitively high in large systems. This leads to the two key questions addressed in the present work:

1. *Can we implement the scheduling technique using the information on flow rates from only a subset of the total network configurations?*
2. *If yes, what is a good subset of network configurations in which the flow rates have to be measured?*

In the following sections, we present arguments substantiating that the scheduling technique could indeed be implemented with information from only a subset of the network configurations, and propose an algorithm for choosing a good subset of network configurations or system states.

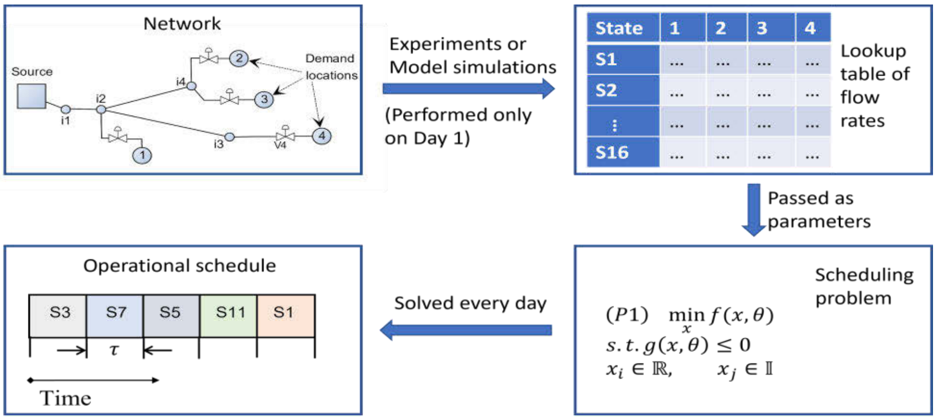

*Figure 2. Schematic describing the scheduling technique (**A1**) presented in* (2).

3 A HEURISTIC FOR STATE DISCOVERY

3.1 Is a subset of network configurations enough?

In algorithm $\mathcal{A}_1$, it is assumed that the lookup table (see *Figure 2*) contains information about all network states. This enables the scheduling problem to prepare an operational schedule that is optimal. In the event the table is incomplete, the program could still prepare a schedule using the network configurations for which data is available. In this case, the resultant schedule need not be optimal as the solution space is only a subset of the original feasible set. If one could ensure that the new search space still contains the optimal solution (or a near optimal solution), then the resultant schedule would still be (near) optimal. In other words, if the information available in the lookup table corresponds to network configurations that are likely to be present in the optimal schedule chosen by the original problem, any schedule prepared using this information can also be expected to be near optimal. In the following paragraphs, we propose a heuristic for identifying such *good* network configurations that could preferably be considered for inclusion in the lookup table.

3.2 Iterative scheme for discovery of states

The algorithm for discovery of states ($\mathscr{A}2$) begins with known flow rates from a few random network configurations. The number of states chosen in this manner is approximately equal to the total number of valves and pumps in the network. With these available states, the scheduling problem can be solved to determine a schedule. However, this schedule would be far from optimum. The next task would be to find a new state that can improve the schedule by the largest extent possible. The flow rates of this new state may be identified and added to the database to prepare the schedule once again. This process of scheduling and state selection may then be repeated until an acceptable schedule is obtained.

In the procedure described above, the challenge is in identifying the next state to be measured/simulated after every iteration. For improvement of the objective function, the new state that is chosen for simulation/measurement should preferentially supply water to the demand points that face the maximum deficiency in supply (negative deviation of supply from demand). This can be achieved by a state that provides water *only* to the demand points with insufficient supply. That is, in the current schedule, if two demand points receive insufficient supply, the newly chosen state should provide water to these two demand points alone. The reason behind the argument is that, in a network configuration supplying *only* to a few demand nodes, these nodes tend to receive a high flow rate, as the others are turned off. This new state can now be simulated/measured, and the flow rates added to the database. Thereafter, the procedure has to be repeated until the resultant schedule is acceptable.

The selection procedure described above could potentially lead to infinite cycles, e.g., the algorithm predicts a new state that has already been selected in a previous iteration. To avoid this, the selection process is made probabilistic. That is, the set of chosen nodes may not be the entire set of nodes receiving insufficient supply in the previous iteration. It may even be a subset of it with certain probability. To implement this, in each iteration, the demand points with insufficient supply have to be selected, their shortfall normalised and then sorted in descending order of the magnitude of the deficit in supply. Thereafter, we pick only a subset of these demand points for supply in the next state. The demand nodes on the top of the list have the maximum deficit in supply and need to be chosen for supply with a high probability while those lower down the list can be chosen with a lower probability, as they are in a much better position with respect to demand satisfaction. We start selecting from the top and go down the list until a certain 'stopping node' is reached. This 'stopping node' has to be decided probabilistically. For this, in each iteration of the heuristic, we generate a random number r between 0 and 1. Demand points are now chosen from the beginning of the sorted list until the sum of their normalised deficit in supply reaches r. If r is close to zero, only the first demand point, i.e., the one with the maximum deficit would be chosen. If r is one, all demand points with shortfall in supply would be chosen. For intermediate values of r, a subset of the nodes with inadequate supply would be chosen. It has to be noted that, for any chosen demand node, any other node with a higher deficit in supply would also be there in the set of chosen nodes. These chosen demand points would receive supply in the new state that has to be measured. The stochastic or probabilistic feature of the algorithm ensures that the algorithm does not get caught in infinite cycles.

The heuristic is further improved by ensuring that a completely new state is chosen randomly, albeit with a small probability. This is included because the rationale used here for choosing new states is ideal only for an unconstrained scheduling problem. If there are constraints in implementation (time or resources), one cannot expect the state selection procedure to be perfect. In the cases we tested, in each iteration, we chose the next state randomly with a probability of 0.05 and with the remaining probability, chose the next state based on the shortfall in supply. In the flow chart of the heuristic given in *Figure 3*, the block verifying the condition - $p >= 0.05?$ - decides whether a state is chosen completely randomly or not. If FALSE (Probability = 0.05), a random state is chosen from the set of unexplored states. Otherwise (TRUE, Probability = 0.95), a state is chosen following the steps described in earlier paragraphs. This procedure is similar to the *ε-greedy exploration* described in (13)

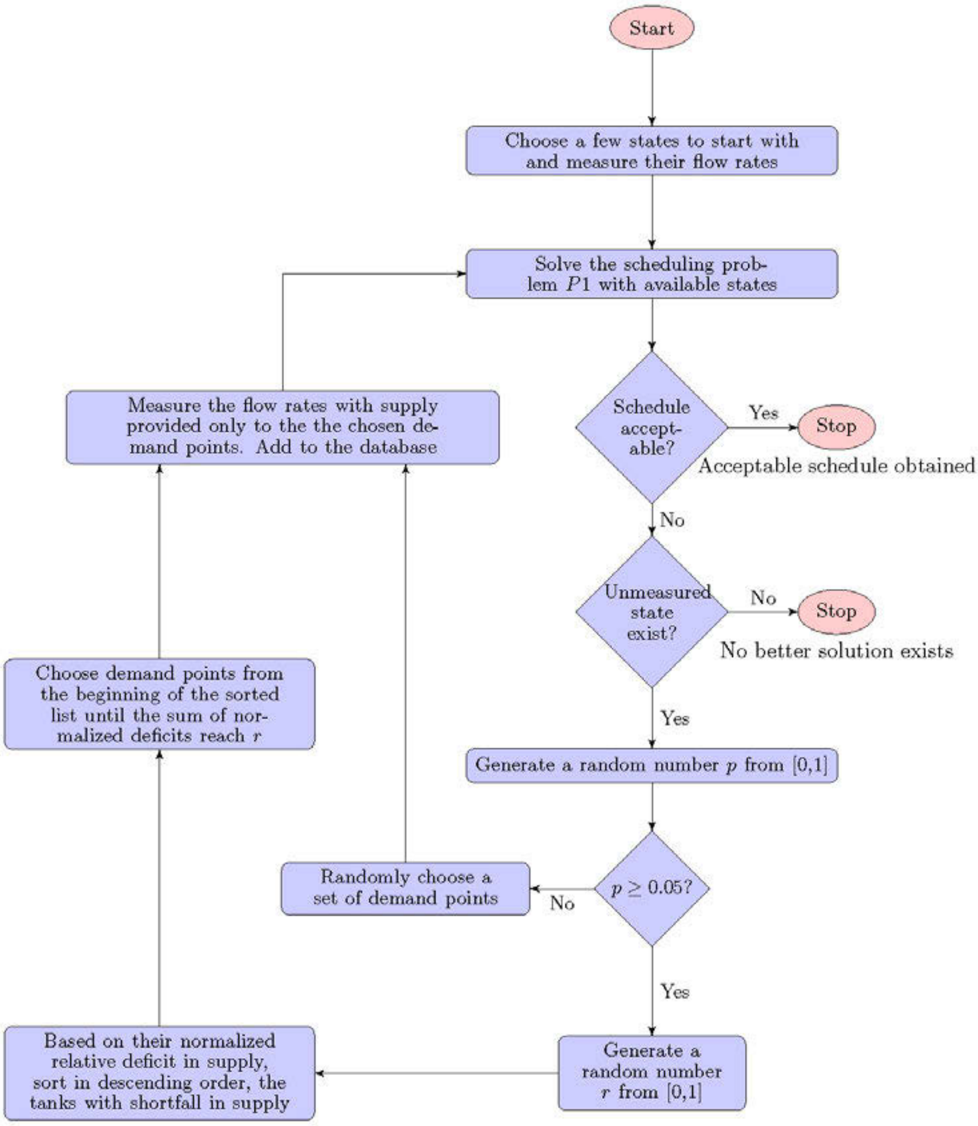

Figure 3 Flowchart describing the algorithm for discovery of states ($\mathcal{A}2$)

In the following sections, we show the implementation of the algorithm in both in-silico and real (laboratory scale) networks.

4 CASE STUDIES – MODEL SIMULATIONS

We tested the heuristic on the mathematical model of the Osmanabad network shown in *Figure 1*. The objective of the scheduling problem was to meet the demand given in *Table 1*. (Please see (2) for more details of the scheduling problem). To start with, it was required to have a set of states for which flow rates were known. Six randomly chosen states, S1 (all valves off), S2, S3, S26, S44 and S64 were used. The states and the flow rates corresponding to each of these are given in *Figure*

4. Thereafter, the heuristic was performed and after each iteration, the flow rates of one new state were identified. This was then added to the database and the procedure was repeated.

Table 1. Eight-hour demand of the villages in Osmanabad

Village	Demand (m^3)
1	81.01
2	30.00
3	30.00
4	52.80
5	75.00
6	50.00

In the first iteration, on preparing the schedule with the six states mentioned above, the sum squared relative deviation (value of the objective function in the scheduling problem) was 2.94. Villages 3 and 5 reported the highest relative deviation of 100% and the deviations corresponding to remaining villages are available in *Table 2*. A random number p was generated to decide whether to choose a state completely random or not. p was greater than 0.05 and hence the next state had to be chosen based on the relative deviations in supply. The random number generated for state selection was 0.8147 (r = 0.8147). As per the procedure described earlier, states had to be selected from the top of the list until the sum of normalised relative deviation reached r. The normalised relative deviation for the first two villages (3 and 5) added up to 0.662 ($< r$) and the first three villages (3, 5 and 6) added up to 0.978 ($\geq r$). Hence, the three villages at the top had to be supplied water in the new state to be explored. This corresponds to state 53 as shown in *Figure 4*. Following this, on preparing a schedule with the seven available states (six initial and the newly explored state S53), the objective function value dropped to 0.45 (from 2.94 in the previous iteration).

Table 2. Villages and their relative deviation on preparing a schedule with the initial six states (sorted on relative deviation)

Village	Relative deviation between supply and demand	Normalised shortfall in supply
1	-1	0.331
2	-1	0.331
3	-0.955	0.316
4	-0065	0.022
5	0.039	NA
6	0.131	NA

Repeating the procedure for five more iterations, states 51, 59, 33, 49 and 8 were explored and the objective function in the final iteration came down to an acceptable level of 0.016. The largest relative deviation for any village had also reduced to 0.09, this time for Village 2. All states chosen by the heuristic for preparing the schedule are shown in *Figure 4* and the variation in objective function value of the optimisation problem with the progression of the heuristic is given in *Figure 5*. It has to be noted that, out of the total 63 non-trivial states, the heuristic required measurements of only 11 states to prepare a reasonable schedule for supply.

	S2	S3	S8	S26	S33	S44	S49	S51	S53	S59	S64
1	18.36		18.36	18.36		18.36					18.36
2		17.23	17.23			15.59		16.60		15.59	15.59
3									11.30		0.00
4				51.58		40.20				47.08	40.20
5							18.81	16.15	17.11	0.00	0.00
6					8.91	5.32	7.32	7.03	7.13	5.54	5.32

Figure 4. The states of the Osmanabad network identified using the heuristic. In the table, each column corresponds to a state and each row corresponds to a village. A colored cell indicate that the valve leading to the village is open in the corresponding state. The numbers in the cell give the flow rate received by the village in m^3 / h.

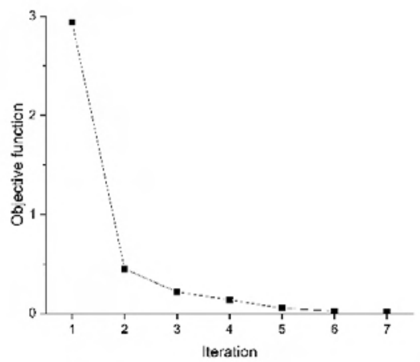

Figure 5. The change in the scheduling objective function (deviation between demand and supply) with the progression of the heuristic.

5 CASE STUDY – EXPERIMENTAL SYSTEM

In this section, we present the results obtained while implementing the method on a laboratory scale WDN.

5.1 Experimental Network:

The network comprises one Over the Head Tank (OHT) supplying water to nine Small Tanks (STs) located downstream. A schematic of the network is shown in *Figure 6* with the STs numbered T1 - T9. The network is an extended version of the setup described in [14] with nine STs instead of five. The pipe and tank dimensions remain unchanged. The control valve installed downstream of OHT was kept open throughout this study. The STs have solenoid valves (Burkert 6011), placed at the inlet and the outlet. The valves could switch ON and OFF the inflow and the outflow(drain) for each ST. Ultrasonic level transmitters (Baumer U500) are installed on top of all tanks to monitor the level. All transmitters and actuators are interfaced to a computer using DAQ card form National Instruments and they are programmed and controlled using LabVIEW. The OHT represent the source and STs represent the storage available at beneficiary villages. The objective is to equitably distribute water by scheduling the solenoid valves which regulates supply to STs in the setup. Two different scenarios are tested to verify the utility of the heuristic for state discovery.

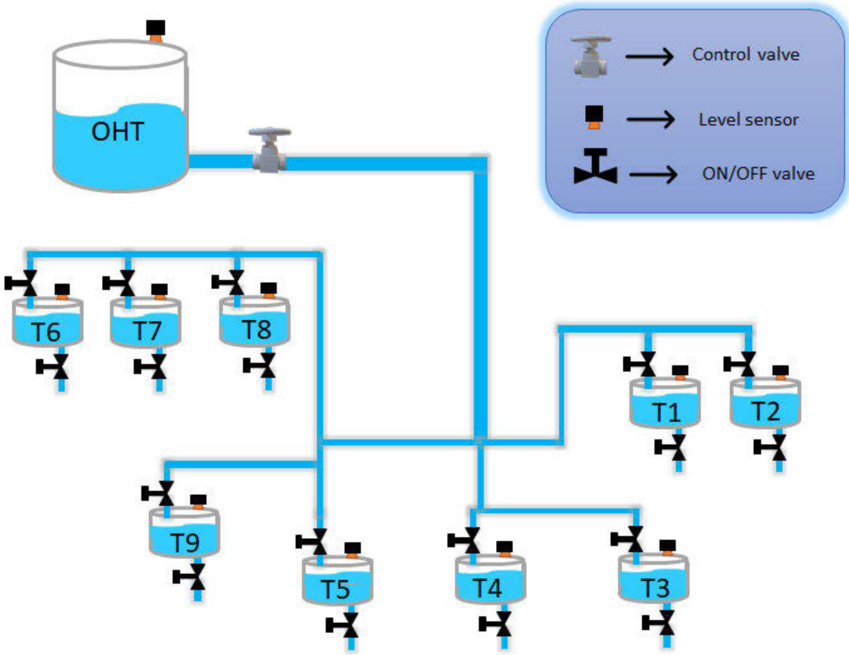

Figure 6. The Experimental Network

5.2 Data Acquisition:

The level of water in the OHT was maintained between 45 cm and 40 cm with a control loop in LabVIEW. The experiments were carried out to identify the inflow rates into STs under different valve combinations. The total number of valve combinations for the system with nine STs is 512(i.e., 2^9). Since it is difficult to measure flow rates for all 512 combinations., we use the heuristic for discovering useful states. We begin with flow rates of few states chosen random and then flow rates of the network configurations chosen by the heuristic was measured and added to database iteratively.

The Ultrasonic level sensor placed at top of the tank gives 4-20 mA analog output. This analog output is sampled at 5Hz frequency and then its filtered and converted to corresponding to level value in LabVIEW. The system was kept idle for 15s before and after the valve activation for every combination, for the level readings to stabilise, following which, level measurements were recorded. Each configuration is kept active for 60s and the difference in the level in each ST was used to calculate the inflow rates into respective ST. The solenoid valve at outflow of STs is kept closed while recording the data.

5.3 Scenario-1

A schedule has to be prepared to meet the demand of 7 litres for each ST. The total time available for water supply is 24 minutes which is discretised into 24 equal intervals of 60 s. The following restrictions are also imposed.

- No of time intervals allowed for the system is 24.
- No of switching allowed for each valve is 4.

We formulated scheduling problems and discovered operational states for the system following the heuristic described in Section 3. To start with, a set of 9 random states were used. Then, the heuristic ($\mathcal{A}2$) was carried out to discover new states and after each iteration within the heuristic, flow rates of one new state were identified.

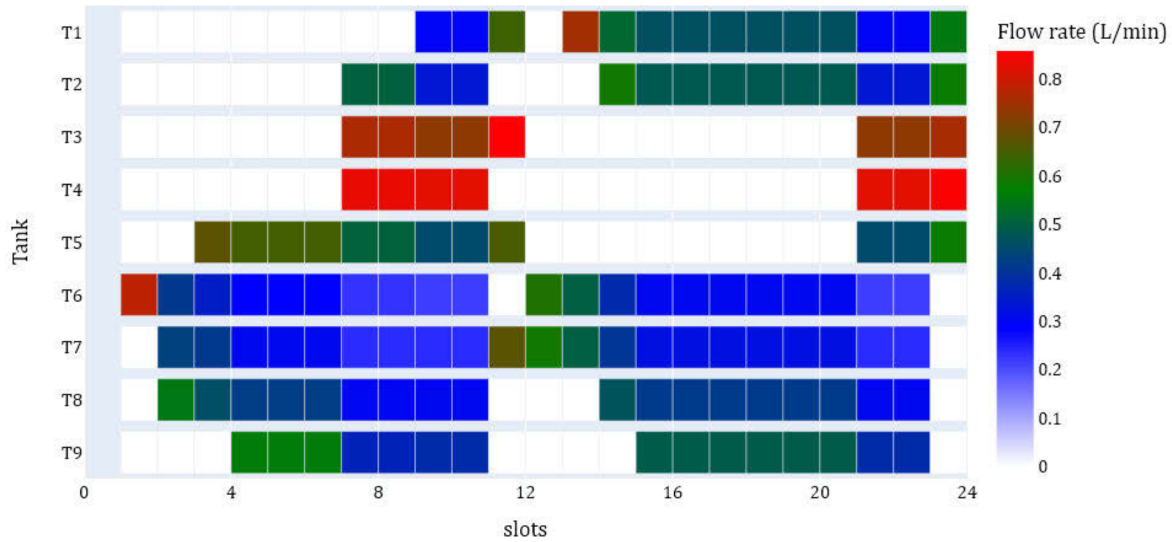

Figure 7. The Final schedule for scenario-1 identified using the heuristic

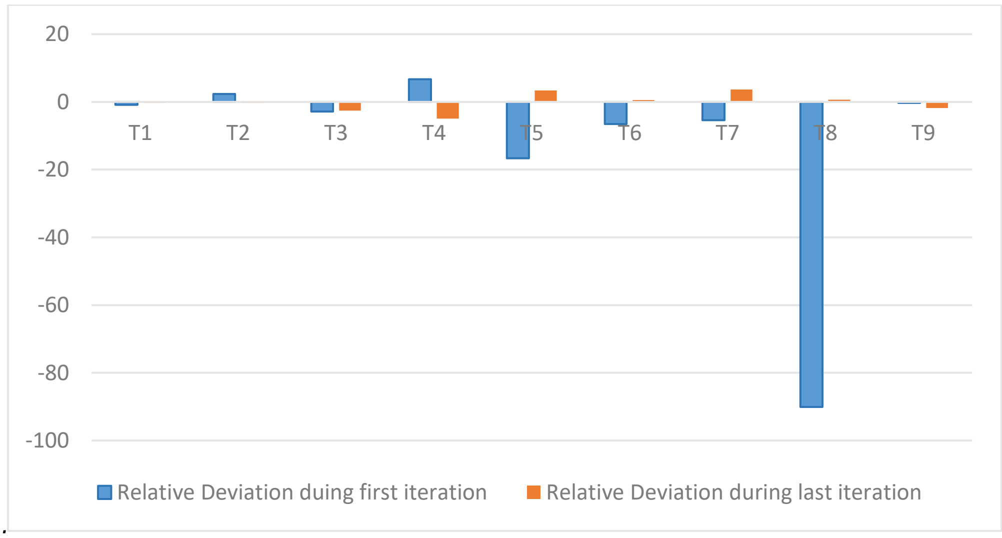

Figure 8. Relative Deviation for scenario-1 in percentage

This was added to the database and the procedure was repeated. The iteration is continued till the maximum relative deviation for every ST was less than 10%. After 7 iterations, a schedule with a maximum relative deviation of 3.6% for tank T7 was obtained. The final schedule and the supply obtained by different STs is shown in the *Figure 7*. In *Figure 7*, each row represents a ST and each column represent a state (network configuration). The coloured cell represents the flow rate leading to that ST is ON and the colour range indicates the flow rate range. The relative deviation in the final schedule is shown in the *Figure 8*.

5.4 Scenario-2

In the second scenario considered here, we added a more stringent condition on the obtained schedule - the total number of switches allowed for each valve was reduced to 2. The scheduler ran initially with 9 different random states and then new states were added with new iterations with the help of the heuristic ($\mathcal{A}2$). The final schedule was obtained after 16 iterations with a maximum relative deviation of 7.4% for T5. The final schedule is shown in the figure 11. On comparing *Figure 7*and *Figure 9*, it is evident that the number of valve operations did decrease substantially in the latter case.

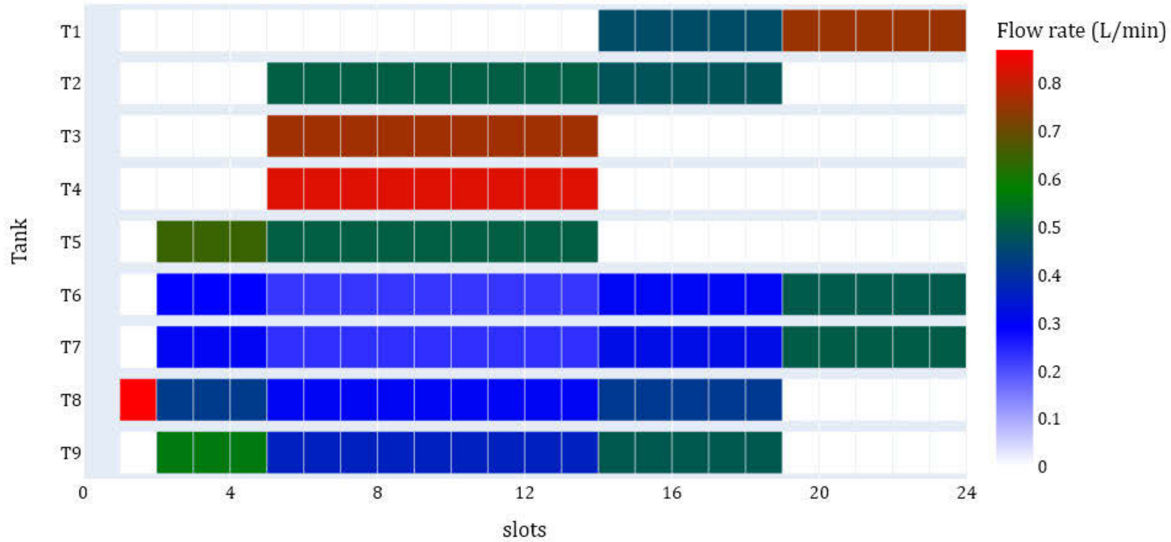

Figure 9. Final schedule for scenario-2 identified using the heuristic.

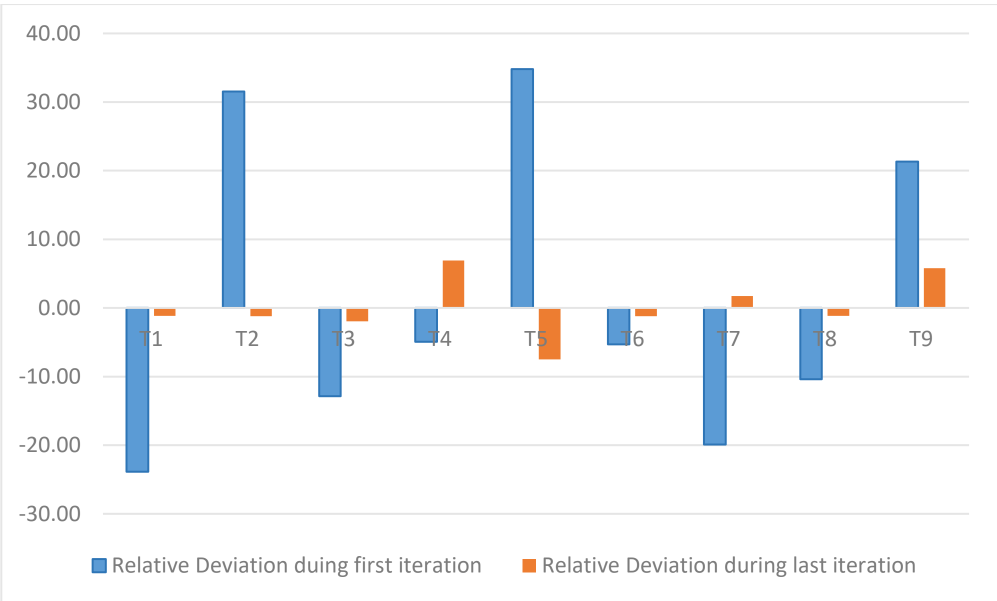

Figure 10. Relative Deviation for scenario-2 in percentage

Out of the total 512 non-trivial states, the heuristic required measurements of only 25 states to prepare a near optimal schedule for supply to STs. It may also be noted that the number of states required in the scenario with more stringent constraints (Scenario 2) was more than that of the case with less stringent constraints (Scenario 1). Through these experiments, we were able to establish that near optimal scheduling with less relative deviation can be obtained with a selective subset of network configurations identified with the help of the heuristic ($\mathcal{A}2$).

6 CONCLUSIONS

In this work, we proposed a heuristic for design of experiments to aid the scheduling in rural WDNs. To the best of our knowledge, this is the first attempt to develop a methodology for design of experiments in WDNs specifically for scheduling. The approach uses cues from a schedule prepared using the available measurements to identify the new network configuration that has to be explored. We demonstrated the applicability of the technique through model simulations and experiments performed on a lab scale WDN. The results show a clear advantage for identifying the states using the heuristic.

The approach presented here assumes that the states are identified prior to the network being commissioned for normal operation. That is, the experimentation does not affect the regular supply of water. The now popular reinforcement learning approaches may help exploring the network states while the network is operational and may also allow correcting the existing measurements for scaling in the pipes. We are also exploring strategies for identifying states that are robust to changes in demands.

7 ACKNOWLEDGEMENTS

This work was partially funded by the Department of Science and Technology, Govt. of India under the Water Technology Initiative (No: DST/TM/WTI/WIC/2K17/82(G) WATER-IC for SUTRAM of EASY WATER) Madras and by the Robert Bosch Center for Data Science and Artificial Intelligence, IIT Madras.

8 REFERENCES

[1] Bonvin, G., S. Demassey, C. Le Pape, N. Maïzi, V. Mazauric, and A. Samperio, “A convex mathematical program for pump scheduling in a class of branched water networks,” Applied Energy, 2018, pp. 1702–1711.

[2] Kurian, V., P. Mohandoss, S. Chandrakesa, S. Chinnusamy, S. Narasimhan and S. Narasimhan, “Optimal scheduling of water distribution networks for equitable supply,” In prep.

[3] Burgschweiger, J., B. Gnadig, and M. C. Steinbach, “Optimization models for operative planning in drinking water networks,” Optimisation and Engineering, vol. 10, no. 1,2009, pp. 43-73.

[4] D’Ambrosio, C., A. Lodi, S. Wiese, and C. Bragalli, “Mathematical programming techniques in water network optimisation,” European Journal of Operational Research, 2010, pp. 10-16.

[5] Mala-Jetmarova, H., N. Sultanova, and D. Savic, “Lost in optimisation of waterdistribution systems? a literature review of system operation,” Environmental Modelling & Software, 2017, pp. 209-254.

[6] Amrutur, B., M. M. Kumar, K. S. Kumar, L. Patel, R. Sundaresan, and N. K. Vaidhiyan, Wateropt, “a method for checking near-feasibility of continuous water supply,” In Proceedings of the 2nd International Workshop on Cyber-Physical Systems for Smart Water Networks. IEEE, 2016.

[7] Kurian, Varghese, Saravanan Chinnusamy, Ashok Natarajan, Sridharakumar Narasimhan, and Shankar Narasimhan. "Optimal operation of water distribution networks with intermediate storage facilities," Computers & Chemical Engineering,2018, pp. 215-227.

[8] Bonvin, Gratien, Sophie Demassey, and Andrea Lodi. "Pump scheduling in drinking water distribution networks with an LP/NLP-based branch and bound." Optimisation and Engineering vol.22, no. 3,2021, pp. 1275-1313.

[9] Prasad, Pooja and Sohoni, Milind. Sugave Water Scheme, Multi-village drinking water scheme analysis. 2011.

[10] Dini, Mehdi, and Massoud Tabesh. "A new method for simultaneous calibration of demand pattern and Hazen-William's coefficients in water distribution systems," Water Resources Management, vol. 28, no. 7, 2014, pp. 2021-2034.

[11] Ostfeld, Avi, Elad Salomons, Lindell Ormsbee, James G. Uber, Christopher M. Bros, Paul Kalungi, Richard Burd et al. "Battle of the water calibration networks," Journal of Water Resources Planning and Management, vol.138, no. 5, 2012, pp. 523-532.

[12] Bhave, Pramod R. and Gupta, Rajesh. Analysis of water distribution networks, 1st ed., Alpha Science Int'l Ltd,2006.

[13] Russo, Daniel J., Benjamin Van Roy, Abbas Kazerouni, Ian Osband, and Zheng Wen, "A tutorial on thompson sampling," Foundations and Trends® in Machine Learning, vol. 11, no.1, 2018, pp. 1-96

[14] Chinnusamy, Saravanan, Prasanna Mohandoss, Varghese Kurian, Sridharakumar Narasimhan, and Shankar Narasimhan, "Operation of intermittent water distribution systems: An experimental study," In Computer Aided Chemical Engineering, vol. 44, 2018, pp. 1975-1980.