IDENTIFICATION OF MONITORING STATIONS IN WATER DISTRIBUTION SYSTEM

By Arun Kumar,[1] M. L. Kansal,[2] and Geeta Arora[3]

ABSTRACT: The location of water quality monitoring stations in a water distribution system should be such that the network is represented with the minimum number of monitoring stations. The present guidelines do not describe how the sampling locations are to be optimally identified. The concept of "pathways" helps in reducing the number of monitoring stations and shows that the best set of stations is that which maximizes the demand coverage of the network. The optimal locations are then identified using integer programming. However, dimensionality of the integer programming increases manifold with the increase in network size and for multiple demand patterns. In this study, modifications in this procedure are suggested to reduce the computational efforts and to make the procedure simpler. The present study identifies the location of monitoring stations from the hydraulics of flow and eliminates the use of cumbersome integer programming. The algorithm is illustrated with the help of two examples.

INTRODUCTION

Providing safe and adequate drinking water for populations in developing countries like India continues to be an area of active research. Massive investments have been made to provide safe drinking water to both urban and rural populations of the country. Various standards are established for maintaining the physical and chemical quality of water [*Indian Manual* (1991)]. Though regulations require specific water quality at the tap, the water regulatory bodies are generally concerned about water quality as it leaves the treatment plant. There is, however, growing awareness that water quality may deteriorate substantially during transport from the treatment plant to the consumer. For this reason, monitoring of water quality at various locations in a public water supply network is needed. However, monitoring of water quality requirements may vary with system size and vulnerability status of various locations. As there is no uniform schedule for monitoring, the degree of variability among monitoring requirements poses both managerial and technical barriers. *The Indian Manual on Water Supply and Treatment* (1991) has prescribed the sampling frequency and water quality parameters to be monitored. Prescribed sampling frequency is based on size of population served by the distribution system, and the sampling is to be spaced evenly over time. Though it is mentioned that the sampling stations are supposed to be representative, yet an adequate procedure for achieving the goal of representative sampling in a distribution system is not specified.

Recently, the procedure of selecting the monitoring stations has drawn wide attention. The most frequently chosen stations are those locations that are in low-demand zones and are more vulnerable from a water-quality point of view. However, in most of the situations, it is not possible to rank the vulnerability of various locations. Thus, in a water supply network, where the demand locations are of unknown vulnerability, some scientific approach is needed to identify the number and location of monitoring stations. One such scientific approach suggested by Lee and Deininger (1992) is based on the concept of demand coverage (DC). The term DC is used to represent the percentage of network demand monitored by a particular monitoring station.

[1]Prof., Civ. Engrg., Delhi Coll. of Engrg., Delhi 110006, India.
[2]Asst. Prof., Civ. Engrg., Delhi Coll. of Engrg., Delhi 110006, India.
[3]Lect., Civ. Engrg., Delhi Coll. of Engrg., Delhi 110006, India.

Note. Associate Editor: J. B. Neethling. Discussion open until January 1, 1998. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on January 19, 1995. This paper is part of the ***Journal of Environmental Engineering***, Vol. 123, No. 8, August, 1997. ©ASCE, ISSN 0733-9372/97/0008-0746–0752/$4.00 + $.50 per page. Paper No. 9974.

Lee and Deininger (1992) have used the information obtained from hydraulic analysis of the network to identify the pathways such that the water quality of a large portion of the network is assessed by installing a few sampling stations. The information obtained from the pathways (in terms of a water fraction matrix) is then converted into an integer programming problem under a chosen coverage criteria. By this method, the lowest level of knowledge occurs when only a very small fraction of the water passes through the node that was called "any fraction." However, for a large network this method becomes highly cumbersome and difficult to handle because of the large dimensionality of the problem. The present study shows that instead of integer programming formulation a far more simple procedure can be used for identifying the locations of monitoring stations. The procedure is based on the hydraulics of flow and is illustrated using a moderately sized network. The results obtained by the suggested procedure have been verified from the study conducted by Lee and Deininger (1992).

SUGGESTED PROCEDURE FOR LOCATION OF MONITORING STATIONS

The problem of water quality monitoring is formulated by assuming that a steady-state hydraulic solution of the water distribution network is available, i.e., all flows and their direction through the network system that satisfy mass balance at the nodes are known and all head losses around closed loops add to zero. There are many hydraulic models [Kentucky Pipe, Gessler and Walski (1985), etc.] available that can analyze a water supply network and can determine the flows with direction in various pipelines and pressures at various nodes. Once the hydraulics of the network is known, one can use the information obtained for identifying the location of various monitoring stations. Strictly speaking, if D_i is the demand at the ith monitoring node and D is the total demand of the system, then, by monitoring the ith node, the water quality of the fraction D_i/D of the network demand is known. Thus, to cover the entire demand of the network, practically every demand node is to be sampled. However, the number of sampling stations can be significantly reduced if the following reasonable assumption is made: The quality of water deteriorates with the passage of time as it flows in the network from the source to various demand nodes. In other words, nodes that are closer to the source (in terms of travel time) receive better quality water than the nodes away from the source.

With this assumption in mind, the steps to be followed for locating the monitoring stations are as follows:

1. For a given demand pattern, analyze the network hy-

J. Environ. Eng. 1997.123:746-752.

draulically and determine the flow and direction of flow in various pipelines.

2. On the basis of the flow directions, renumber the nodes (if required) in such a way so that the water flows from lower to higher node number.
3. Compute the elements of the water fraction matrix (starting from lowest numbered node to highest) as follows:
   a. Choose one node of the water distribution network.
   b. Determine all nodes connected to the chosen node.
   c. Calculate water fractions of all the upstream nodes connected to the chosen node. Here, upstream nodes are those nodes that are closer to the source and water flowing through these enroute to the monitoring station. Similarly, downstream nodes are those nodes where the flowing out water of the monitored node will eventually reach. A water fraction of an individual upstream node is the flow of an individual upstream node contributing to the chosen node divided by the total inflow of the chosen node. (A detailed procedure for estimating the water fraction matrix is given in a subsequent section).
4. Choose an appropriate coverage criteria. For example, in the present study, a 60% criteria is chosen. Here, 60% criteria means that a particular upstream node is assumed to be monitored by monitoring water quality at a chosen node if 60% or more of the inflow to the chosen node has passed through this particular upstream node.
5. From the water fraction matrix computed in step 3, deduce the coverage matrix as follows: If water fraction values computed in Step 3 are greater than the chosen criteria (i.e. ≥ 0.6), it can be assumed that the particular upstream nodes are covered by monitoring at the chosen node and are assigned a value of 1. If the calculated water fraction value of any upstream node is less than the chosen criteria, the particular upstream node is not covered by monitoring at the chosen node and is assigned a value of 0.
6. Select each demand node as the monitoring station. Compute the DC for all such selected nodes in the coverage matrix by adding the demands of all those upstream nodes that are marked as '1' and are enroute to the chosen node.
7. Choose that node as the first monitoring station that has the maximum DC as computed in step 6. Also, identify the nodes covered by this monitoring station. (These will be the nodes whose demands have been added to compute the DC of the monitoring station.)
8. For locating the subsequent monitoring station, eliminate the row corresponding to the already chosen monitoring station from the coverage matrix. Also, eliminate the columns corresponding to the nodes covered by the already chosen monitoring station.
9. Again compute the DC of the remaining nodes and choose that node as the next monitoring station that now has the maximum DC.

In this suggested procedure, most of the computational efforts are involved in the computation of the water fraction matrix. Once the water fraction matrix is ready, it is easier to deduce the coverage matrix that depends upon the coverage criteria chosen. From the coverage matrix, one can easily find the first and subsequent monitoring station by way of computing the DC. In the subsequent sections, the computation of the water fraction matrix, coverage matrix, and DC has been explained in detail. The whole methodology is illustrated with the help of an example.

Computation of Water Fraction Matrix [W]

For calculating the demand coverage by a monitoring node, the heuristic presented here makes use of the concept of water fraction matrix $[W]$ introduced by Lee and Deininger (1992). The demand nodes are renumbered such that water flows from a lower to a higher numbered node. This renumbering facilitates the computations and also makes the $[W]$ matrix a lower triangular matrix that represents the water fraction of the upstream nodes connecting the source to sampling node. The water fraction matrix $[W]$ can then be represented as

$$[W] = \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ \vdots \\ W_n \end{bmatrix} = \begin{bmatrix} w_{11} & & & & \\ w_{12} & w_{22} & & & \\ w_{13} & w_{23} & w_{33} & & \\ \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot & \\ w_{1n} & w_{2n} & w_{3n} & \cdot\cdot & w_{nn} \end{bmatrix} \quad (1)$$

where w_{ij} = fraction of demand d_j at the jth monitored node that has passed through upstream ith node; $\mathbf{W}_j$ = row vector of elements w_{ij} ($1 \leq i \leq j$) that represents the fraction of demand D_j at the jth monitoring node that has passed through some or all of i upstream nodes. The nonzero value of w_{ij} in $\mathbf{W}_j$ vector represents those i nodes that are in the path of water reaching the jth node; $w_{ij} = 0$ means that no water from node i is reaching node j; and $w_{jj} = 1$ means that whole of demand D_j has passed through jth node.

Start with the lowest numbered node vector $\mathbf{W}_1$. For this node, there will be only one element (w_{11}) which will be '1' since all the demand drawn at this node (D_1) has passed through itself only. The process of computation of w_{ij} should be carried out in an ascending node number order. For any node, all the upstream nodes that feed water directly to this particular node should be identified. For example, let a node numbered n be fed directly by three upstream nodes k, l, and m. If Q_{kn}, Q_{ln}, and Q_{mn} are the quantities of flow and P_{kn}, P_{ln}, and P_{mn} are the respective proportion of water from nodes k, l, and m to the node numbered n, the proportion of water received directly from these nodes will be as follows:

$$P_{kn} = \frac{Q_{kn}}{Q_{kn} + Q_{ln} + Q_{mn}} \quad (2a)$$

$$P_{ln} = \frac{Q_{ln}}{Q_{kn} + Q_{ln} + Q_{mn}} \quad (2b)$$

$$\text{and } P_{mn} = \frac{Q_{mn}}{Q_{kn} + Q_{ln} + Q_{mn}} \quad (2c)$$

Now, these directly feeding nodes k, l, and m are also getting supply from further upstream nodes 1, 2, 3.... The water fraction of all these upstream nodes to the sampling node numbered n will be vector $\mathbf{W}_n$ which may be given as

$$\mathbf{W}_n = (w_{1n}, w_{2n}, \ldots, w_{kn}, w_{ln}, w_{mn}, \ldots, w_{n-1,n}, w_{nn}) \quad (3)$$

The elements of $\mathbf{W}_n$ can be computed as

$$\mathbf{W}_n = [(P_{kn} \times \mathbf{W}_k + P_{ln} \times \mathbf{W}_l + P_{mn} \times \mathbf{W}_m), w_{nn}] \quad (4)$$

$$\mathbf{W}_n = [P_{kn} \times (w_{1k}, w_{2k}, \ldots, w_{kk}) + P_{ln} \times (w_{1l}, w_{2l}, \ldots, w_{ll}) + P_{mn} \times (w_{1m}, w_{2m}, \ldots, w_{mm}), 1] \quad (5)$$

Rearranging (5) and comparing with (3)

$$w_{1n} = (P_{kn} \times w_{1k} + P_{ln} \times w_{1l} + P_{mn} \times w_{1m})$$
$$w_{2n} = (P_{kn} \times w_{2k} + P_{ln} \times w_{2l} + P_{mn} \times w_{2m})$$
$$w_{3n} = (P_{kn} \times w_{3k} + P_{ln} \times w_{3l} + P_{mn} \times w_{3m})$$
$$\vdots$$
$$\vdots$$
$$w_{n-1,n} = (P_{kn} \times w_{n-1,k} + P_{ln} \times w_{n-1,l} + P_{mn} \times w_{n-1,m})$$

and

J. Environ. Eng. 1997.123:746-752.

$$w_{nn} = 1$$

These elements of $\mathbf{W}_n$ are written if three nodes (k, l, and m) are feeding directly to the sampled node n. However, on similar lines, one can compute the various vector components for any other number of directly feeding nodes to the sampled node.

Computation of Demand Coverage (DC)

A percentage proportion contribution should be fixed as governing criteria for coverage of an upstream node by monitoring water quality at a downstream node. This chosen percentage proportion should represent the adequate criteria for a particular node being covered by monitoring water quality of other node. One of this criteria may be that the node is assumed to be covered if majority of the sampled water at chosen node has come from that particular upstream node. The logic is based on the fact that the water collected at the jth station has partly or wholly passed from some of the upstream demand nodes and if the quality at the downstream monitored node j is good, there is every likelihood that the water quality of the upstream set of i demand nodes also must be good. This is particularly valid as the downstream monitored node receives delayed water and water quality deteriorate (TTHMs, bacteria and chlorine residuals, etc.) with time and distance from the source. In the present study, assume that if 60% or more of the inflow to a particular monitoring station has passed through an upstream node, then it is reasonable to say that the water quality at this upstream node can be inferred from the quality of the monitored node. Therefore, the water fraction matrix prepared earlier is modified to a 60% coverage area matrix in which all values of $w_{i,j}$ that are ≥ 0.6 are changed to one and all the remaining values of $w_{i,j}$ to 0.

Once a suitable percentage criteria is chosen, the demand coverage of any node can then be calculated as

$$\mathrm{DC}_i = (w_{1i} \quad w_{2i} \quad w_{3i} \ldots w_{ii}) \begin{pmatrix} D_1 \\ D_2 \\ D_3 \\ \vdots \\ Di \end{pmatrix} \tag{6}$$

The node that gives maximum value of demand coverage will be the first monitoring station.

Subsequent monitoring stations can be located by computing the DC of all the remaining nodes in the coverage matrix after eliminating the row of the chosen monitoring station and the columns corresponding to those upstream nodes that have been covered indirectly by the chosen monitoring station.

ILLUSTRATIVE EXAMPLE

A simple water distribution network (as shown in Fig. 1) has been considered, which has 19 delivery nodes and 34 pipe links with two sources of supply. It is assumed that the quality of water for both the sources is identical and there is complete mixing of flows from these two sources. The network is hydraulically analyzed using the popular WADISO program given by Gessler and Walski (1985). The flow in various pipelines and the corresponding flow directions with renumbered nodes have been shown in Fig. 1. The water fraction matrix [**W**] is generated considering each node (starting from lower to higher) as the sampling node. The elements w_{ij} of water fraction matrix [**W**] are calculated and are shown in Table 1. Some representative calculations are presented in the following:

- At node 1, there is no upstream demand node through which the water reaches at node 1. Hence all P_{ij} elements are zero. Using (4), $\mathbf{W}_1 = w_{1,1}$ and will be 1.

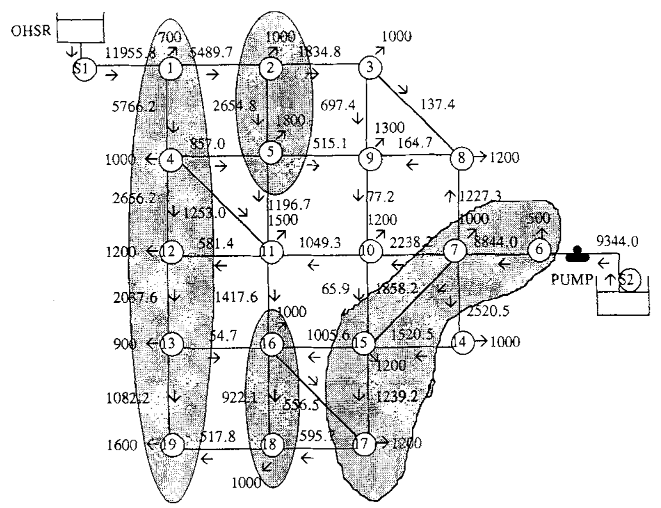

FIG. 1. Example Network of Water Distribution System

TABLE 1. Water Fraction Matrix for Example Network of Fig. 1

Sample at node (1)	CONTRIBUTING NODES (m^3/d) 1 700 (2)	2 1,000 (3)	3 1,000 (4)	4 1,000 (5)	5 1,800 (6)	6 500 (7)	7 1,000 (8)	8 1,200 (9)	9 1,300 (10)	10 1,200 (11)	11 1,500 (12)	12 1,200 (13)	13 900 (14)	14 1,000 (15)	15 1,200 (16)	16 1,000 (17)	17 1,200 (18)	18 1,000 (19)	19 1,600 (20)
1	1																		
2	1	1																	
3	1	1	1																
4	1	0	0	1															
5	1	0.756	0	0.244	1														
6	0	0	0	0	0	1													
7	0	0	0	0	0	1	1												
8	0.101	0.101	0.101	0	0	0.899	0.899	1											
9	0.892	0.801	0.518	0.091	0.374	0.108	0.108	0.120	1										
10	0.029	0.026	0.017	0.003	0.012	0.971	0.971	0.004	0.033	1									
11	0.709	0.267	0.005	0.442	0.346	0.291	0.291	0.001	0.010	0.300	1								
12	0.948	0.048	0.001	0.900	0.062	0.052	0.052	0′	0.002	0.054	0.180	1							
13	0.948	0.048	0.001	0.900	0.062	0.052	0.052	0	0.002	0.054	0.180	1	1						
14	0	0	0	0	0	1	1	0	0	0	0	0	0	1					
15	0.001	0	0	0	0	0.998	0.998	0	0.001	0.019	0	0	0	0.441	1				
16	0.427	0.154	0.003	0.273	0.199	0.573	0.573	0.001	0.006	0.181	0.576	0.022	0.022	0.179	0.406	1			
17	0.133	0.049	0.001	0.085	0.062	0.867	0.867	0	0.003	0.069	0.179	0.007	0.007	0.359	0.816	0.310	1		
18	0.312	0.113	0.002	0.199	0.145	0.688	0.688	0.001	0.005	0.137	0.420	0.016	0.016	0.250	0.567	0.730	0.392	1	
19	0.742	0.069	0.002	0.672	0.089	0.258	0.258	0	0.003	0.081	0.258	0.681	0.681	0.081	0.184	0.237	0.127	0.324	1

Note: Numbers 1–19 in columns 2–20 represent the numbers of the contributing nodes; and values 700–1,600 in these columns are the demands in m^3/d at these contributing nodes.

TABLE 2. Coverage Matrix Under 60% Criteria for Example Network of Fig. 1

Sample at node (1)	CONTRIBUTING NODES (m^3/d)																			Demand covered (m^3/d) (21)
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
	700 (2)	1,000 (3)	1,000 (4)	1,000 (5)	1,800 (6)	500 (7)	1,000 (8)	1,200 (9)	1,300 (10)	1,200 (11)	1,500 (12)	1,200 (13)	900 (14)	1,000 (15)	1,200 (16)	1,000 (17)	1,200 (18)	1,000 (19)	1,600 (20)	
1	1																			700
2	1	1																		1,700
3	1	1	1																	2,700
4	1	0	0	1																1,700
5	1	1	0	0	1															3,500
6	0	0	0	0	0	1														500
7	0	0	0	0	0	1	1													1,500
8	0	0	0	0	0	1	1	1												2,700
9	1	1	0	0	0	0	0	0	1											3,000
10	0	0	0	0	0	1	1	0	0	1										2,700
11	1	0	0	0	0	0	0	0	0	0	1									2,200
12	1	0	0	1	0	0	0	0	0	0	0	1								1,900
13	1	0	0	1	0	0	0	0	0	0	0	1	1							2,800
14	0	0	0	0	0	1	1	0	0	0	0	0	0	1						2,500
15	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1					2,700
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1				1,000
17	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	1			3,900
18	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	1		3,500
19	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	5,400

Note: Numbers 1–19 in columns 2–20 represent the numbers of the contributing nodes; and values 700–1,600 in these columns are the demands in m^3/d at these contributing nodes.

TABLE 3. Modified Coverage Matrix for 2nd Monitoring Station

Sample at node (1)	CONTRIBUTING NODES (m^3/d)														Demand covered (cmd) (16)
	2	3	5	6	7	8	9	10	11	14	15	16	17	18	
	1,000 (2)	1,000 (3)	1,800 (4)	500 (5)	1,000 (6)	1,200 (7)	1,300 (8)	1,200 (9)	1,500 (10)	1,000 (11)	1,200 (12)	1,000 (13)	1,200 (14)	1,000 (15)	
1															0
2	1														1,000
3	1	1													2,000
4	0	0													1,000
5	1	0	1												2,800
6	0	0	0	1											500
7	0	0	0	1	1										1,500
8	0	0	0	1	1	1									2,700
9	1	0	0	0	0	0	1								2,300
10	0	0	0	1	1	0	0	1							2,700
11	0	0	0	0	0	0	0	0	1						1,500
12	0	0	0	0	0	0	0	0	0						1,200
13	0	0	0	0	0	0	0	0	0						2,100
14	0	0	0	1	1	0	0	0	0	1					2,500
15	0	0	0	1	1	0	0	0	0	0	1				2,700
16	0	0	0	0	0	0	0	0	0	0	0	1			1,000
17	0	0	0	1	1	0	0	0	0	0	1	0	1		3,900
18	0	0	0	1	1	0	0	0	0	0	0	1	0	1	3,500

Note: Numbers 2–18 in columns 2–15 represent the numbers of the contributing nodes; and values 1,000–1,000 in these columns are the demands in m^3/d at these contributing nodes.

- At node 2, the entire water has been directly fed by node 1 only. Hence $P_{1,2} = 1.0$ and $w_{1,2} = P_{1,2} \times \mathbf{W}_1$; $w_{22} = 1$. Thus, $\mathbf{W}_2 = [1.0 \times (1), 1] = (1, 1)$
- Similarly, at node 5, the direct feeding is by node 2 and 4 only. So, $P_{2,5} = 2{,}654.8/(2{,}654.8 + 857) = 0.756$ and $P_{4,5} = 857/(2{,}654.8 + 857) = 0.244$.
- The vector $\mathbf{W}_5$ can be computed by using (4) as

$$\mathbf{W}_5 = (P_{2,5} \times \mathbf{W}_2 + P_{4,5} \times \mathbf{W}_4),\ w_{5,5}$$

$$\mathbf{W}_5 = [0.756 \times (1, 1) + 0.244 \times (1, 0, 0, 1)],\ w_{5,5}$$

$$\mathbf{W}_5 = [(0.756 + 0.244, 0.756 + 0, 0, 0.244)],\ 1$$

$$\mathbf{W}_5 = (1, 0.756, 0, 0.244, 1)$$

- At node 17

$$\mathbf{W}_{17} = P_{15,17} \times \mathbf{W}_{15} + P_{16,17} \times \mathbf{W}_{16},\ w_{17,17}$$

$$\mathbf{W}_{17} = 0.690 \times (0.001, 0, 0, 0, 0, 0.998, 0.998, 0, 0.001, 0.019, 0, 0, 0, 0.441, 1) + 0.310 \times (0.427, 0.154, 0.003, 0.273, 0.199, 0.573, 0.573, 0.001, 0.006, 0.181, 0.576, 0.022, 0.022, 0.179, 0.406, 1),\ 1$$

$$\mathbf{W}_{17} = (0.133, 0.049, 0.001, 0.085, 0.062, 0.867, 0.867, 0, 0.003, 0.069, 0.179, 0.007, 0.007, 0.359, 0.816, 0.310, 1)$$

After generating the water fraction matrix, a 60% coverage area matrix is derived from it, as shown in Table 2. All those values of $w_{i,j}$, which are ≥ 0.6 are assigned values of 1 and others 0. The demand coverage (DC_i) values are calculated for each of the 19 nodes using (6) and the node with the maximum value of DC_i is chosen as the first monitoring station. From the column of DC_i in Table 2, it is clear that node 19 will be the first monitoring station, as it has a maximum demand cov-

TABLE 4. Monitoring Stations

Monitoring station(s) (1)	Demand covered (cmd) (2)	Nodes covered (3)
19	5,400	1, 4, 12, 13, 19
17	3,900	6, 7, 15, 17
5	2,800	2, 5
18	2,000	16, 18

erage of 5,400 m^3/d out of a total 21,300 m^3/d and the nodes covered by this monitoring station will be 1, 4, 12, 13, and 19.

For locating the second monitoring station, the row corresponding to monitoring station 19 and columns corresponding to nodes 1, 4, 12, 13, and 19 are removed from the coverage matrix given in Table 2 and the DC for all the remaining nodes are calculated again as shown in Table 3. Node 17 will be the next optimal monitoring station as it has now the maximum value of DC_i' in Table 3. The demand covered by node 17 is 3,900 m^3/d and the nodes covered by this monitoring station are 6, 7, 15, and 17. Similarly, one can find the subsequent monitoring stations. Some of these monitoring stations, their demand coverage, and the nodes covered by them are listed in Table 4.

It may be observed from Table 1 that more than 60% of water received at node 19 has passed via nodes 1, 4, 12, 13, and 19. By monitoring water quality at node 19, definite quality inference (for minimum 60% quality assurance) can be drawn about water quality at nodes 1, 4, 12, and 13 in addition to node 19.

If one has to choose only four monitoring stations in the example mentioned earlier, the best four will be 19, 17, 5, and 18 having a demand coverage of 14,100 m^3/d (5,400 + 3,900 + 2,800 + 2,000) out of a total of 21,300. However, by monitoring water quality at these four sampling stations, no quality statement can be given about water withdrawn from the uncovered nodes (3, 8, 9, 10, 11, 14). If any of these uncovered nodes lies in the high vulnerable zone, one should identify the same and choose that as an additional monitoring station.

To verify the results obtained from the suggested methodology, the same problem has been solved using the integer programming model suggested by Lee and Deininger (1992) in Appendix I. It can be observed that the four monitoring stations and the various demand nodes covered by the two methodologies are exactly the same.

VALIDATION OF METHODOLOGY

The central thrust of the present study is to avoid duplication of monitoring in the network as well as to identify the low-contributing nodes that left uncovered after selecting a particular set of monitoring stations from the hydraulics of flow. The selection of monitoring stations is based on the hypothesis that a downstream node will have poorer water quality as compared to corresponding upstream nodes and if the water quality at downstream node is good then the water quality of all the respective upstream demand nodes is also good. This has been obtained by way of identifying a set of upstream and downstream nodes from the hydraulic analysis of the network. One can also find the location of monitoring stations with the help of the water quality concentration model. However, the procedure will be cumbersome and difficult as the water quality concentration parameters are dynamic in nature and involve complex reaction equations. To verify the poorer water quality of the monitoring node (among the nodes for which it is representative), a simplified water quality concentration model named as EPANET (Rossman 1994) is used. The EPANET algorithm is based on the Discrete Volume Element Method (DEVM) (Rossmann et al. 1993), which provides a mechanism for considering the loss (or growth) of a substance by reaction as it travels through the distribution network. It has been assumed that a first-order decay reaction takes place within the bulk flow and with material along the pipe wall. Within each hydraulic time period when flows are constant, DVEM computes a shorter water quality time step and divides each pipe into a number of completely mixed volume segments. Within each water quality time step, the material contained in each pipe segment is first transferred to its adjacent downstream segment. When the adjacent segment is a junction node, the mass and flow entering the node is added to any mass already received from other pipes. After this transport step is completed for all pipes, the resulting mixture concentration at each junction is computed and released into head end segments of pipes with flow leaving the node. Then the mass within each pipe is reacted. This sequence of steps is repeated until the time when a new hydraulic condition occurs. The network is then resegmented and the computations are continued.

For illustration purposes, water quality in terms of residual chlorine content at various demand nodes has been computed. The general expression for the reaction rate of chlorine decay in bulk flow is taken as

$$\theta(c) = -Kc \tag{7}$$

where $\theta(c)$ = rate of chlorine decay (mg/L/day); $K = k_b + \{k_w k_f[R_H(k_w + k_f)]\}$; k_b = bulk rate (day^{-1}); k_w = wall reaction constant (per day); k_f = mass-transfer coefficient (m/s); R_H = hydraulic radius of pipe (m); and c = initial concentration in bulk flow (mg/L).

For computing the residual chlorine at various demand nodes using the EPANET model, the initial chlorine content at all the sources has been taken as 12 mg/L and at various nodes as 0.0 mg/L. Minimum travel time from one node to another is taken as 6 min and chemical diffusivity as 1.30 e^{-8} m^2/s. The coefficient of bulk and wall decay has been assumed to be 0.5 per day. Using these initial parameters, the steady-state chlorine contents at various nodes were found and are tabulated in Table 5.

From Table 5, it is clear that node 19 (which is representative of nodes 1, 4, 12, 13, and 19 as per the suggested procedure) has the lowest chlorine content among the nodes 1, 4,

TABLE 5. Residual Chlorine Content at Various Demand Nodes of Network Shown in Fig. 1

Node number (1)	Chlorine content (mg/L) (2)
1	10.76
2	9.57
3	8.49
4	9.56
5	8.31
6	12.00
7	10.60
8	8.41
9	5.73
10	8.87
11	7.04
12	7.78
13	6.88
14	9.33
15	8.53
16	6.40
17	6.03
18	4.90
19	4.97
S1	12.00
S2	12.00

J. Environ. Eng. 1997.123:746-752.

TABLE 6. Water Fraction Matrix for Example Network of Fig. 2 with Flow Scenario 1 and 2 [after Lee and Deininger (1992)]

Sample at node (1)	CONTRIBUTING NODES (Demand in m³/d) 1 (2)	2 (3)	3 15 (10)* (4)	4 30 (10)* (5)	5 25 (5)* (6)	6 20 (45)* (7)	7 10 (30)* (8)	Demand Covered under Flow Scenarios 1 (9)	2 (10)	Total demand covered (11)
1	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)			
2	1 (1)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)			
3	1 (1)	1 (1)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	15	10	12.5
4	1 (1)	1 (1)	1 (1)	1 (1)	0 (0)	0 (0)	0 (0)	45	20	32.5
5	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	0 (0)	0 (0)	70	25	47.5
6	1 (1)	1 (1)	0 (1)	0 (1)	0 (1)	1 (1)	1 (0)	30	70	50
7	1 (1)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	1 (1)	10	30	20

Note: Values in bracket are for flow scenario 2, otherwise for 1; and numbers 1–7 in columns 2–8 represent the numbers of the contributing nodes. *These values are the demands at the contributing nodes for flow scenarios one and two (without parentheses pertains to flow scenario one, and in parentheses represents flow scenario two).

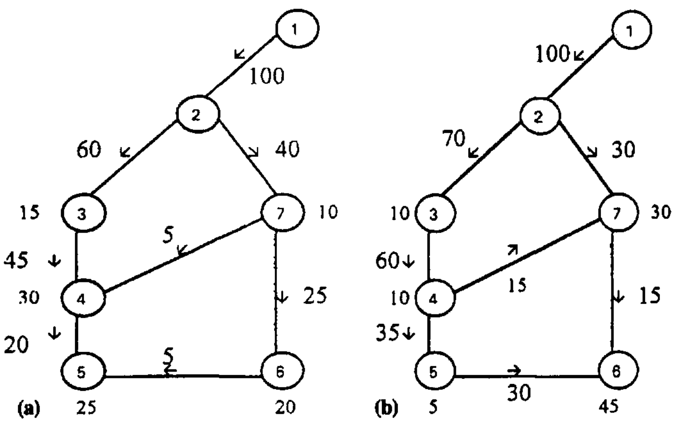

FIG. 2. Simple Water Distribution System: (a) Flow Scenario 1; (b) Flow Scenario 2 [after Lee and Deininger (1992)]

12, 13, and 19. Thus, if the substance is conservative in nature, node 19 will be representative of nodes 1, 4, 12, 13, and 19. Similarly, node 17 (which is representative of nodes 6, 7, 15, and 17) has the lowest chlorine content among nodes 6, 7, 15, and 17 and is thus representative of water quality for nodes 6, 7, 15, 17, and so on. This verifies the suggested procedure for identifying the location of monitoring stations.

MULTIPLE FLOW SCENARIOS

In real network, resource demands vary hourly, daily, and seasonally. Thus while locating the monitoring stations one has to keep in mind the various demand patterns and different flow scenarios. These varying demand patterns can be taken into account by formulating the corresponding water fraction matrix for any sampling node i. For a given network, if there are n flow scenarios that exist for time periods $x_1, x_2, \ldots, x_n$ respectively, (6), for computing the demand coverage of any chosen node, would be modified accordingly. DC_i^1, DC_i^2, $\ldots$, DC_i^n are the demand coverage values for flow scenarios 1, 2, 3, . . . , n, respectively, then the average demand coverage of the ith monitoring station, during the entire time period is given as

$$DC_i = \frac{X_1 \times DC_i^1 + X_2 \times DC_i^2 + \cdots + X_n \times DC_i^n}{x_1 + x_2 + \cdots + x_n} \quad (8)$$

The multiple demand scenarios have been considered in the example network as shown in Fig. 2. The example that is taken from Lee et al. (1992) is a water distribution network with seven nodes out of which five are demand nodes. The flows in various pipelines and nodal withdrawals for both the demand patterns are shown in the Fig. 2. Also, it has been considered that each demand pattern and hence the flow scenario exists 50% of time. The coverage matrix has been generated from the water fraction matrices for the two flow scenarios under a 50% coverage criterion. The demand coverage (DC_i) by various nodes as monitoring stations has been computed using (8) and are tabulated in Table 6.

For example, if node 3 is the sampling node, the demand coverage (DC_3) is calculated using (8) as shown in the following:

As $x_1 = x_2 = 0.5$; $DC_3^1 = 15$ and $DC_3^2 = 10$, so, $DC_3 = (0.5 \times 15 + 0.5 \times 10)/1 = 12.5$.

Similarly, $DC_4 = (0.5 \times 45 + 0.5 \times 20)/1 = 32.5$, and $DC_5 = 47.5$; $DC_6 = 50$; $DC_7 = 20$.

It can be observed that node 6 has the maximum demand coverage of 50 units under 50% coverage criterion. Thus, node 6 should be the first optimal station that is the same as that suggested by Lee and Deininger (1992). One can choose the different demand coverage criteria depending upon the local conditions and the resources available. In that case, the values of the water fraction matrix elements will change, which in turn may change the values in demand coverage matrix. However, the procedure of identifying the location of monitoring station(s) will remain the same.

CONCLUSIONS

The objective of this paper is to identify the optimal location of monitoring stations in a water distribution network from the hydraulics of flow. This has been obtained by way of avoiding duplication or repetition of monitoring in the network and by maximizing the demand coverage in the network. After selecting a set of monitoring stations, the suggested procedure is to identify the nodes that are low-contributors to the downstream nodes and are not covered by any of the selected monitoring stations. To monitor water quality of all the demand nodes, additional monitoring is required at these uncovered nodes. Previously, this optimization problem has been solved using integer programming that takes into account the various constraints for a given node pattern of the water distribution system. The integer programming problem accommodates $(n + 1)$ constraints and $(n \times p + n)$ variables, where n is the number of delivery nodes in the distribution system and p is the number of demand patterns. For a large network, the dimensions of the problem become very large. Moreover, such programming efforts, though simple for most academic purposes, find less applicability for a practicing engineer. Simpler and less specialized methods, though unverifiable, are of greater interest to a nonacademician. This paper suggests a simplified procedure for identifying the optimal locations of monitoring stations in a water distribution system. The procedure is illustrated through two examples demonstrating the superiority of the proposed procedure.

J. Environ. Eng. 1997.123:746-752.

ACKNOWLEDGMENTS

The writers sincerely acknowledge the guidance given by P. B. Sharma, Principal, Delhi College of Engineering, Delhi, during the course of this study. The writers are also grateful to the anonymous reviewers for their critical review and constructive suggestions to improve the quality and readability of this paper.

APPENDIX I. FORMULATION AS INTEGER PROGRAMMING OPTIMIZATION PROBLEM [AFTER LEE AND DEININGER (1992)]

Given the water distribution network shown in Fig. 1 and a coverage criterion of 60%, the question arises of where four monitoring stations should be located to maximize the demand covered.

Two sets of integer variables are introduced. The x_i ($i = 1, \ldots, n$) set denotes whether there is a sampling station at node i, and the y_i ($i = 1, \ldots, n$) set denotes whether the demand at node i is covered. Both sets of integer variables are 0, 1 variables.

Mathematically, the problem is

$$\text{maximize} \sum_{i=1}^{n} d_i y_i \qquad (9)$$

subject to

$$\sum x_i \leq \text{NS}; \quad \sum \bar{a}_{i,j} x_i - y_i \geq 0 \qquad (10, 11)$$

where, x_i, y_i = 0, 1; NS = number of sampling stations; d_i = demand at node i; $a_{i,j}$ = transpose of the coverage matrix; and n = number of nodes.

For the illustrative example shown in Fig. 1, the problem formulation is as follows:

$$\begin{aligned}\text{Maximize } & 700y_1 + 1{,}000y_2 + 1{,}000y_3 + 1{,}000y_4 \\ & + 1{,}800y_5 + 500y_6 + 1{,}000y_7 + 1{,}200y_8 + 1{,}300y_9 \\ & + 1{,}200y_{10} + 1{,}500y_{11} + 1{,}200y_{12} + 900y_{13} + 1{,}000y_{14} \\ & + 1{,}200y_{15} + 1{,}000y_{16} + 1{,}200y_{17} + 1{,}000y_{18} + 1{,}600y_{19}\end{aligned} \qquad (12)$$

subject to

$$\begin{aligned}& x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} \\ & + x_{12} + x_{13} + x_{14} + x_{15} + x_{16} + x_{17} + x_{18} + x_{19} \leq 4\end{aligned} \qquad (13)$$

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_9 + x_{11} + x_{12} + x_{13} + x_{19} - y_1 \geq 0 \qquad (14)$$

$$x_2 + x_3 + x_5 + x_9 - y_2 \geq 0; \quad x_3 - y_3 \geq 0 \qquad (15, 16)$$

$$x_4 + x_{12} + x_{13} + x_{19} - y_4 \geq 0; \quad x_5 - y_5 \geq 0 \qquad (17, 18)$$

$$x_6 + x_7 + x_8 + x_{10} + x_{14} + x_{15} + x_{17} + x_{18} - y_6 \geq 0 \qquad (19)$$

$$x_7 + x_8 + x_{10} + x_{14} + x_{15} + x_{17} + x_{18} - y_7 \geq 0 \qquad (20)$$

$$x_8 - y_8 \geq 0; \quad x_9 - y_9 \geq 0; \quad x_{10} - y_{10} \geq 0 \qquad (21\text{–}23)$$

$$x_{11} - y_{11} \geq 0; \quad x_{12} + x_{13} + x_{19} - y_{12} \geq 0 \qquad (24, 25)$$

$$x_{13} + x_{19} - y_{13} \geq 0; \quad x_{14} - y_{14} \geq 0 \qquad (26, 27)$$

$$x_{15} + x_{17} - y_{15} \geq 0; \quad x_{16} + x_{18} - y_{16} \geq 0 \qquad (28, 29)$$

$$x_{17} - y_{17} \geq 0; \quad x_{18} - y_{18} \geq 0; \quad x_{19} - y_{19} \geq 0 \qquad (30\text{–}32)$$

where x_{ij} and y_{ij} are integer variables.

Solution of this problem gives the values of x_5, x_7, x_{18}, and x_{19}, as 1, meaning, therefore, that the four monitoring stations should be installed at nodes 5, 7, 18, and 19. Similarly, the values of variables y_1, y_2, y_4, y_5, y_6, y_7, y_{12}, y_{13}, y_{15}, y_{16}, y_{17}, y_{18}, and y_{19} are 1, meaning that the demand nodes covered by these four monitoring stations will be 1, 2, 4, 5, 6, 7, 12, 13, 15, 16, 17, 18, and 19. The total demand covered by these four monitoring stations will be 14,100 cmd (i.e., the summation of demands of the covered demand nodes).

APPENDIX II. REFERENCES

Gessler, J., and Walski, T. M. (1985). "WADISO—Water distribution system optimization." *Tech. Rep. EL-85-11, Waterways Experiment Station*, U.S. Corps of Engineers, Vicksburg, Miss.

Indian manual on water supply and treatment. (1991). Ministry of Urban Development, New Delhi, India.

Lee, B. H., and Deininger, R. A. (1992). "Optimal locations of monitoring stations in water distribution system." *J. Envir. Engrg.*, ASCE, 118(1), 4–16.

Lee, B. H., Deininger, R. A., and Clark, R. M. (1991). "Locating monitoring stations in water distribution systems." *J. AWWA*, 83(7), 60–66.

Rossman, L. A. (1994). *EPANET—Users manual.* United States Environmental Protection Agency (EPA), Cincinnati, Ohio.

Rossman, L. A., Boulos, P. F., and Altman, T. (1993). "Discrete volume-element method for network water-quality models." *J. Water Resour. Plng. and Mgmt.*, ASCE, 119(5), 505–517.

APPENDIX III. NOTATION

The following symbols are used in this paper:

- D_j = demand withdrawn from node j of network (m^3/d);
- DC_i = demand coverage, defined as total quantity of water about quality of which definite inference can be drawn by selecting node i as monitoring node (m^3/d);
- DC_i^a = demand coverage of node i calculated for particular demand scenario a (m^3/d);
- DC_m = maximum demand coverage (m^3/d);
- $P_{k,n}$ = ratio of water received at node n that has passed via node k to total water received at node n;
- $Q_{k,n}$ = quantity of water flowing from node k to n (m^3/d);
- $[W]$ = water fraction matrix, elements of which are $w_{i,j}$;
- $\mathbf{W}_j$ = row vector of water fraction matrix; elements of this row are contributions of node i to j towards water reaching node j;
- $w_{i,j}$ = fraction of water received at node j that has come via node i;
- $\mathbf{w}'_{i,j}$ = modified water fraction matrix element (modification is made every time after selecting one of the nodes as sampling node);
- $w_{m,j}$ = elements of $\mathbf{W}_m$; and
- X_a = time duration of flow scenario a in hours.

J. Environ. Eng. 1997.123:746-752.