LEAK LOCATION IN WATER DISTRIBUTION NETWORKS BASED ON DYNAMIC TESTS AND PARAMETRIC IDENTIFICATION

F. Casella*, L. Bascetta*, C. Maffezzoni*, G. Bodini

 * Politecnico di Milano, Dipartimento di Elettronica e Informazione, Piazza Leonardo da Vinci, 32, 20133, Milano, Italy Tel. +39 02 2399 3400 - Fax. +39 02 2399 3412 {maffezzo, casella, bascetta}@elet.polimi.it

Keywords: Dynamic tests, Fault detection and isolation, Hydraulic networks, Leak location, Parameter estimation, Statistical testing.

Abstract

The paper deals with the problem of locating leaks in water distribution networks with many nodes and branches. The pressure transients measured in a few selected nodes, in response to pressure variations applied to one node, are analyzed by a detection and isolation algorithm, based on parametric estimation of faulty network models. The algorithm is tested by simulation in a realistic case, giving promising results.

1 Introduction

Water leakage is a costly problem, both because of wasting a precious natural resource and in economic terms. The primary economic loss concerns the cost of raw water, its treatment and transportation. Moreover leakage ineluctably causes damage to the pipe network itself and to the foundations of roads and buildings. These and other reasons have increased, in recent years, the interest towards the study of efficient methods for leak location in underground or otherwise not accessible piping.

Traditionally, this task is accomplished using simple and direct methods based on stationary state calculations, such as flow balance methods [2], on acoustic measurements [11] or on direct physical inspection. However, these methods suffer from the drawbacks of being sensitive to uncertain parameters, and/or to require invasive measurements and possibly excavation works. As an alternative, methods based on dynamic tests, where the modifications to the wave propagation patterns in the pipe, due to the presence of leaks, are studied, appeared in the literature in the early 1980's. The typical application for these methods is the location of one or more leaks in a single underwater or underground pipeline, by analyzing pressure and, possibly, flow rate transients at the pipe extremes.

Different approaches are adopted. For instance, Brunone [3] proposes a very simple dynamic test technique to locate existing leaks in an outfall pipe, by analyzing the pressure transient in a suitable position when the head at one end of the

pipe is suddenly increased. More sophisticated methods have been also proposed, using fluid dynamics theory and Fault Detection and Isolation (FDI) techniques [6], based on state observers [1, 13, 10, 14] or on artificial neural networks and fuzzy logic [15], which are able to detect and locate incipient leaks. However, these methods often require flow rate measurements, which are usually unavailable in practice, and suffer from high sensitivity to uncertainty. Moreover these approaches usually deal with the problem of leak detection in a single pipeline.

In this work, the problem of locating existing leaks in a pipe network is dealt with. Brunone's ideas cannot be applied in this case, since the contributions to the pressure signal due to small waves reflected by the leak cannot be spotted in the complex transient pattern generated by a network. The application of observer-based FDI techniques, such as the robust observer approach of Ge and Fang [7], and the Fault Detection Filter initially proposed by White [16] has also been investigated. However, it has been found that these techniques do not scale well to the high-order systems which result from the discretization of even moderate-sized networks (requiring hundreds of state variables). The cited examples in the literature deal with pipe models comprising a few segments. A different method has then been devised, based on parameter estimation, another class of model-based FDI techniques as stated in [8], and on some heuristic assumptions; it is interesting to note that only pressure measurements are required, thus making the method attractive in urban contexts, where the installation of flow meters can be hard. Simulation tests on a realistic sub-urban scale distribution network have been conducted, leading to promising results.

The paper is organized as follows. In Section 2 the modelling of water distribution networks is discussed. The proposed leak location methodology is presented in Section 3 and 4. Simulation results in a realistic case are shown in Section 5. Finally, the conclusions are given in Section 6.

2 Network modelling

Urban water distribution network can be modelled as the connection of pipes, valves, and orifices (which can model either user outlets or pipe leaks). When small-sized portions of the urban network are considered, no components with lumped mass storage are present, such as reservoirs or buffer tanks; therefore, pressure and flow dynamics are entirely due to wave propagation phenomena through the pipes. The simulator employed to generate test data sets can thus describe the static and dynamic behavior of the following components: long pipes, short pipes, valves and outlets.

Pressure and flow dynamics in long pipes are described by the distributed mass and momentum conservation equations [17], neglecting the kinetic term

$$\frac{\partial H}{\partial t} + \frac{c^2}{Ag} \frac{\partial q}{\partial x} = 0$$
$$\frac{\partial q}{\partial t} + Ag \frac{\partial H}{\partial x} + K_f q |q| = 0$$

where H is the head, c the speed of sound, A the pipe crosssectional area, g the gravitational acceleration, K_f a suitable friction factor and q is the volumetric flow rate. Now, consider the pipe as composed of N segments of length Δx , and apply the method of characteristics; under the mild assumption that the flow rate can be considered uniform along a single segment, for the purpose of computing the friction head loss term, the following difference equations are obtained [4]

$$H_B(k) = H_A(k-N) - Z(q_B(k) - q_A(k-N)) + -NK_f \Delta x q_A(k-N) |q_A(k-N)| \quad (1)$$

$$H_A(k) = H_B(k - N) + Z(q_A(k) - q_B(k - N)) + NK_f \Delta x q_B(k - N) |q_B(k - N)|$$
(2)

where the A and B subscripts refer to the head and tail of the pipe, respectively and Z = c/Ag is the characteristic impedance of the pipe. Each step corresponds to a real time delay $\Delta t = c\Delta x$.

The other components are described by simple algebraic equations. In particular, short pipes, whose length is such that the wave propagation delay can be neglected, can be modelled as

$$H_B = H_A - K_f Lq|q|$$

where L is the pipe length. Valves can be modelled by equation of the type

$$q = K_v \sqrt{\rho g (H_A - H_B)} \qquad H_A > H_B$$

where the constant K_v can be obtained from standard formulae, such as those found in [9]. Outlets and leaks can be described by the following equation

$$q = K_l A_l \sqrt{2g(H_A - z_A)}$$

where K_l is a discharge coefficient (usually around 0.8), A_l is the cross-sectional area of the leak and z_A is the node elevation.

Finally, the network equations are obtained by assembling the branch equations with a balance of flows at each node.

An input-output Linear Time Invariant (LTI) model of the network dynamics can be derived from the previous one. As will be explained later, the basic assumption of the location method is that the only active outlets are the leaks; this implies that the actual flows will be much smaller than the nominal ones, and the head losses, being proportional to the squared flow, will be negligible. This assumption has been verified by simulation of realistic scenarios. Moreover, at such low flow rates, the friction term in (1) and (2) is no longer accurate, since the fluid is characterized by a very low Reynolds number.

The LTI model is built according to the following steps:

- 1. the steady-state conditions around which the leak location experiment will be carried out is computed
- 2. the short pipes are eliminated, since they do not contribute significantly to head losses, nor to the dynamics
- 3. each long pipe is split into N connected single-segment pipes, so that equations (1) and (2) involve only terms at steps k and k 1; the friction terms are neglected, thus making the equations linear
- 4. open valves are eliminated from the network, considering the associated head loss negligible, and their terminal nodes are collapsed; closed valves are eliminated and a null flow boundary condition is imposed to the connected pipes. The reason for doing this is that valves in distribution networks are not used with regulation purposes, but only for network segmentation, i.e. any valve can be assumed either fully open or fully closed
- 5. outlet equations are linearized around the operating point found in step 1.

Moreover, the segment number rounding approximation introduced above can be quite crude when obtaining LTI models for estimation purpose, whose order must be kept limited by selecting a sufficiently large segment length Δx . To overcome this problem, following concept borrowed from [5], fractional delays can be introduced in the single-segment pipe model

$$H_B(k) = (1+\alpha)H_A(k-1) - \alpha H_A(k) + - Z(q_B(k) - (1+\alpha)q_A(k-1) - \alpha q_A(k))$$

$$H_A(k) = (1 + \alpha)H_B(k - 1) - \alpha H_B(k) + + Z(q_A(k) - (1 + \alpha)q_B(k - 1) - \alpha q_B(k))$$

where α is equal to $1 - \frac{L}{N\Delta x}$. Note that this approximation is good only within a limited frequency range, i.e. for frequencies up to around $0.1/\Delta t \ Hz$, and can be thus applied to pipe models provided that a suitable low-pass filtering of input/output data is applied.

3 Leak location problem formulation

In the scenario of model-based FDI methods two main classes can be identified [8]: the techniques that rely on the concept of analytical redundancy and the class of methods that relies on parametric estimation. In particular, parameter estimation is a natural approach to residual generation in case of parametric multiplicative faults and can take advantage of a wide knowledge based on system identification theory.

The method here proposed, that can be classified as a parameter estimation FDI techniques, can be applied to an isolated subnetwork, with a single feed where the pressure variations are applied. The network is excited by applying PRBS (Pseudo Random Binary Signal) head variations, which can be generated by connecting the feeding node with two pressurized tanks via a three-way rapid solenoid valve. The portion of the network under consideration must be physically isolated, using valves, or virtually isolated, by measuring the heads at the boundary nodes, from the whole distribution network.

Three main assumptions are made:

- the leaks, one or more, are the only relevant outlets in the network, i.e. all the user outlets are closed (this is a realistic assumption if urban residential areas are considered during late night time);
- only time invariant leaks are considered, i.e. the case of incipient leaks is not taken into account;
- the actual flows, during the experiment, are much smaller than the nominal ones and the head losses can thus be neglected.

The head transients are measured in one or more pipe connection joints, corresponding to manholes.

A pipe network can be represented as a graph in which each node stands for a pipe segment boundary. Given a network graph, its set of nodes Π can be divided into three subsets: the subset Π of the boundary nodes (of cardinality $\tilde{\pi}$), i.e. nodes located at the boundaries of the network or subnetwork under consideration, in which a pressure measurement is set; the subset Π (of cardinality $\hat{\pi}$) of the measurement nodes, i.e. all the nodes in which a pressure measurement is set, not included in the previous subset, and the subset Π^* (of cardinality π^*) of all other nodes without measurements.

Considering now this partition of the set of nodes, the linearized equations of the network in the frequency domain can be expressed, using the balance of flows at each node, as follows

$$\begin{bmatrix} G_{11}(q^{-1}) & G_{12}(q^{-1}) & G_{13}(q^{-1}) \\ G_{21}(q^{-1}) & G_{22}(q^{-1}) & G_{23}(q^{-1}) \\ G_{31}(q^{-1}) & G_{32}(q^{-1}) & G_{33}(q^{-1}) \end{bmatrix} \begin{bmatrix} \tilde{p}_k \\ \hat{p}_k \\ p_k^* \end{bmatrix} = \begin{bmatrix} \tilde{w}_k \\ \hat{w}_k \\ w_k^* \end{bmatrix}$$
(3)

where $G(q^{-1})$ is a matrix of transfer functions, signal vectors $\tilde{w}_k, \hat{w}_k, w_k^*$ are the variations, from their nominal values, of the water flows through the boundaries of the network and through

leakages located in nodes belonging to subsets $\hat{\Pi}$ and Π^* , respectively, and \tilde{p}_k , \hat{p}_k , p_k^* are the variations, from their nominal values, of pressures at each node.

The two subsets of flows \hat{w}_k and w_k^* are related to leaks as follows

$$\hat{w}_k = K_{22}\,\hat{p}_k\tag{4}$$

$$w_k^* = K_{33} \, p_k^* \tag{5}$$

where $K_{22} = K_{22}(\vartheta)$ and $K_{33} = K_{33}(\vartheta)$ are diagonal matrices, functions of the unknown vector ϑ , i.e. the vector of cross-sectional areas of the leaks.

Combining now the matricial equation of the network (3) with the equations (4) and (5), and discarding the first $\tilde{\pi}$ rows, the following equality is derived

$$\tilde{Y}(q^{-1},\vartheta)\tilde{p}_k + \hat{Y}(q^{-1},\vartheta)\hat{p}_k = 0$$

where

$$\hat{Y}(q^{-1},\vartheta) = G_{21}(q^{-1}) + G_{23}(q^{-1}) \left(G_{33}(q^{-1}) - K_{33}\right)^{-1} G_{31}(q^{-1})$$

$$\hat{Y}(q^{-1},\vartheta) = G_{22}(q^{-1}) - K_{22} + G_{23}(q^{-1}) \left(G_{33}(q^{-1}) - K_{33}\right)^{-1} G_{32}(q^{-1})$$

Obviously this equality holds only if the exact areas of the leaks are known. Therefore the leak location problem can be formulated as a nonlinear regression problem, i.e. as the problem of finding the best estimate of the vector ϑ that minimizes the following residual

$$\varepsilon_k(\vartheta) = \tilde{Y}(q^{-1},\vartheta)\tilde{p}_k + \hat{Y}(q^{-1},\vartheta)\hat{p}_k \tag{6}$$

This problem may then be solved using the least squares approach and the best estimate $\bar{\vartheta}$ may be given as follows¹

$$\bar{\vartheta} = \arg\min \left\|\varepsilon_k(\vartheta)\right\|_2 \qquad s.t. \ \vartheta \ge 0$$
(7)

The residual vector $\varepsilon_k(\vartheta)$ may be efficiently evaluated, avoiding numerical problems related to the calculation of the transfer functions $\tilde{Y}(q^{-1},\vartheta)$ and $\hat{Y}(q^{-1},\vartheta)$, by simulating the LTI models of particular partitions of the original network with a specific set of boundary conditions. In fact, considering equations (3) and (6), each residual can be interpreted as the error in the balance of flows of the correspondent node belonging to $\hat{\Pi}$.

Assume now the measurement nodes $\hat{\Pi}$ are chosen so that the network, deprived of these nodes, is partitioned in π_s disjunctive subnetworks $S_1, S_2, \ldots, S_{\pi_s}$, i.e. $\bigcap_{i=1}^{\pi_s} S_i = \hat{\Pi}$, and the subset $\tilde{\Pi}$ includes only one node, i.e. the input node for the

¹Note that the estimation problem here considered is nonlinear in the vector of parameters ϑ .



Figure 1: An example network

pressure perturbation. Consider, as an example, a network (Fig. 1) composed of four subnetworks $(S_1, S_2, S_3 \text{ and } S_4)$, partitioned by three measurement nodes $(\hat{s}_1, \hat{s}_2 \text{ and } \hat{s}_3)$, and with a feeding node (\tilde{s}) . The residual correspondent, for example, to node \hat{s}_1 can be calculated as the sum of flows coming from subnetworks S_1 and S_2 , i.e. from the subnetworks connected to that node, minus the flow through the leak located in \hat{s}_1 . Each of these flows is obtained simulating the LTI model of the correspondent subnetwork, with the current value of the vector ϑ , having the pressure measurements as inputs. For example, the water flow coming from subnetwork S_1 can be evaluated simulating the LTI model of S_1 having, as inputs, the pressure variations $\hat{p}_{1_k}, \hat{p}_{2_k}$ at nodes \hat{s}_1, \hat{s}_2 . Finally, the residuals are obtained, exploiting the superposition principle, as the flow balances at each node $\hat{\Pi}$.

The practical implementation of this algorithm is based on *Matlab* function *lsqnonlin*, that solves the nonlinear least squares problem. Starting from an initial guess, e.g. one leak of zero area for each node of the network, this function finds a minimum to the sum of squares of the residuals, subjected to the positiveness constraint previously introduced. Moreover, the vector of residuals is calculated, at each iteration, simulating the LTI subnetwork models, with the current value of the leakage vector, as previously described.

4 Interpretation of the estimate via hypothesis testing

Let $\bar{\vartheta}$ be a solution of the optimization problem (7) and H_0 : $A\vartheta = 0$ the hypothesis under test², where A is a known $\delta \times (\hat{\pi} + \pi^*)$ matrix of rank δ . The aim of this test is to verify the trustfulness of the hypothesis that a leakage is located among the subsets of nodes defined by the matrix A. Let RSS and RSS_H be the sum of squares of the residuals of the least squares problem (7) and of the same problem including the constraints $A\vartheta = 0$, respectively. Assuming that the residuals are i.i.d. gaussian variables³, it can be proved that [12], when H_0 is true, the quantity

$$F = \frac{(RSS_H - RSS)/\delta}{RSS/[\eta - (\tau - \kappa)]}$$

is distributed as $F_{\delta,\eta-\tau+\kappa}$ (the *F*-distribution with δ and $\eta - \tau + \kappa$ degrees of freedom), where η is the number of samples of the signal $\varepsilon_k(\vartheta)$ considered in the optimization problem (7), τ is the rank of the jacobian matrix associated to the linearized optimization problem and κ is the number of active constraints $(\bar{\vartheta}_i = 0)$ associated to the solution $\bar{\vartheta}$. Chosen then a level of significance σ , a threshold F_{σ} can be calculated from the *F*-distribution and the hypothesis is rejected if $F > F_{\sigma}$.

The method just described can be effectively used to interpret the leak estimate making up an iterative algorithm composed of the following steps:

- 1. let $\Theta = \hat{\Pi} \bigcup \Pi^*$ the set of nodes in which a leak may be located;
- 2. solve the optimization problem (7);
- find the subset Θ_a ⊂ Θ of nodes that belong to an active constraint and set Θ' = Θ \ Θ_a;
- 4. apply the hypothesis testing to all the subsets composed of a single node drawn from Θ' and let Θ_{H_1} be the subset of nodes that reject the hypothesis;
- 5. apply the hypothesis testing to all the subsets composed of two adjacent nodes drawn from $\Theta' \setminus \Theta_{H_1}$ and let Θ_{H_2} be the subset of nodes that reject the hypothesis;
- apply the hypothesis testing to all the subsets composed of three adjacent nodes drawn from Θ' \ (Θ_{H1} ∪ Θ_{H2}) and let Θ_{H3} be the subset of nodes that reject the hypothesis;
- 7. let $\Theta' = \Theta_{H_1} \bigcup \Theta_{H_2} \bigcup \Theta_{H_3}$, where the union must be done in such a way that all the subsets that have not a void intersection are merged together, be the new set of nodes in which a leak may be located;
- 8. if $\Theta' \neq \Theta$ set $\Theta = \Theta'$ and restart from step 2.

5 A simulation case study

A real network, installed in the waste water treatment plant of Cremona (Italy), is here considered. The network spans approximately 45.000 m^2 , with a radius of 300 m and is characterized by both meshed and linear sections, built with steel pipes having diameters between 25 and 80 mm. A sketch of the network structure is shown in Fig. 2. The excitation signal is a PRBS signal, 10s long, filtered by a first order low-pass transfer function emulating the finite valve travel time. Two leaks of 10 mm^2 cross-sectional area (Fig. 2) are simulated.

²Note that the results here presented for the hypothesis testing problem hold only in the case of linear regression. Thus, as the estimation problem here considered is nonlinear, all the consideration that follows are based on a linearized estimation problem around the current solution $\bar{\vartheta}$.

³In fact, as the whiteness hypothesis does not rigorously hold, a technique to whiten the residuals, before doing the hypothesis testing, is currently under study.



Figure 2: Sketch of the network

The feeding node and the measurement nodes have been chosen heuristically as follows (Fig. 2):

- node 1 as the feeding node, since it seems to be almost equally far from all the terminal nodes of the network;
- nodes 4, 5, 7, 13 as the measurement nodes, since this is one of the possible choices to split the network in disjunctive subnetworks.

Making this choice the network is subdivided into seven subnetworks (Fig. 3) balancing a trade-off between the minimization of the number of sensors used and the aim of having low order LTI subnetwork models.

It is very important to test the robustness of the estimation algorithm with respect to various kinds of uncertainty. In particular the following items have been considered:

- variations of the wave propagation velocity from its nominal value, along a pipe or among different zones of the network;
- presence of unmodelled pipe cross section variations, due to pipe fittings or incrustations;
- measurement noise;
- presence of lumped (i.e. open valves) and distributed (i.e. pipes) head losses.

To this aim, the pressure data has been generated with the nonlinear simulator with a sampling time of 1 ms, including measurement noise, unmodelled pipe cross-section variations, and friction effects according to Colebrook-White equations; also, the actual wave propagation velocity has been set to a value which is 2% greater than the nominal one. The sampling time



Figure 3: Subnetwork partitiong

of the LTI models used for estimation is 25 ms, which is a reasonable compromise between accuracy and computational burden. It has been found that the performance of the location algorithm is severely impaired by the wave propagation velocity mismatch, while being tolerant towards the other kinds of uncertainty. Therefore, the wave propagation velocity has been included among the parameters to be estimated. Also, it has been found that high-pass filtering the data which is fed into the least squares algorithm improves the results.

The algorithm presented appears promising: as can be seen from Fig. 4 (at the top of each bar the estimated leak cross-sectional area and the F value found in the correspondent hypothesis test, normalized with the threshold value used and expressed in logarithmic scale, are shown) the two leaks are cor-



Figure 4: Leak location on a real network

rectly found and no sham leak appears. In particular, these leaks are located in the nodes nearest to the real leak position (indicated with an arrow in Fig. 4), as the LTI estimation model considered is spatially discretized in a finite number of segments. To obtain this result two iterations are required: after the first one only 3 of the 44 initial elements of vector ϑ do not belong to active constraints, i.e. the correspondent Lagrange multipliers are equal to zero. Since the second optimization problem gives the same result the algorithm is stopped and an hypothesis testing with a level of significance of the 95% is performed: only two single nodes, correspondent to the real leaks, reject the hypothesis.

6 Conclusions

A methodology for leak location in water pipeline networks, based on dynamic tests and parametric estimation, has been discussed in this paper. The method proposed has been devised to overcome some of the limitations affecting the approaches, based on dynamic tests, that have been developed in the literature, that is the ability of locating more than one leakage, not only in a single pipe but also in a complex meshed network, with a method that can be scaled up to significant sized networks.

A test on a network case study, based on simulation, has been presented. A particular effort has been devoted to check the robustness of the method against all the typical uncertainties that must be taken into account in the framework of a real water pipeline network. Moreover an experimental validation has already been planned as a future work that will be done on the network shown in Section 5.

Finally, a *Matlab* package to perform modelling, simulation, analysis and leak location, has been developed.

Acknowledgments

This work has been supported by CNR, Istituto di Elettronica e Ingegneria dell'Informazione e delle Telecomunicazioni, Sezione di Milano.

References

- L. Billman and R. Isermann. Leak detection methods for pipelines. *Proceedings 8th IFAC Congress*, pages 381– 385, 1984.
- [2] H.H. Brons and H. Schaffhaussen. European methods of leak detection and location. *Pipe Line Industry*, pages 50–66, May 1972.
- [3] B. Brunone. Transient test-based technique for leak detection in outfall pipes. *Journal of Water Resources Plan*ning and Management, 125(5):302–306, 1999.
- [4] G. Ferretti, C. Maffezzoni, and V. Rossi. A novel approach to speed control of hydro power stations. *Automatica*, 26(3):557–565, 1990.

- [5] G. Ferretti, C. Maffezzoni, and R. Scattolini. Recursive estimation of time delay in sampled systems. *Automatica*, 27(4):653–661, 1991.
- [6] P.M. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy - a survey and some new results. *Automatica*, 26(3):459–474, 1990.
- [7] W. Ge and G.Z. Fang. Detection of faulty components via robust observation. *International Journal of Control*, 47:581–599, 1988.
- [8] J.J. Gertler. Fault detection and diagnosis in engineering systems. Marcel Dekker, Inc., 1998.
- [9] ISA. *Flow equations for sizing control valves*. Standard ANSI/ISA-S75.01. 1986.
- [10] Z. Kowalczuk and K. Gunawickrama. Leak detection and isolation for transmission pipelines via nonlinear state estimation. *4th IFAC SAFEPROCESS*, 2:943–948, 2000.
- [11] J.R. De Read. Comparison between ultrasonic and magnetic flux pigs for pipeline inspection. *Pipes and Pipelines International*, pages 7–15, January-February 1987.
- [12] G.A.F. Seber. *Linear regression analysis*. John Wiley and Sons, USA, 1977.
- [13] D.N. Shields, S.A. Ashton, and S. Daley. Desing of nonlinear observers for detecting faults in hydraulic subsea pipelines. *Control Engineering Practice*, 9:297–311, 2001.
- [14] C. Verde. Multi-leak detection and isolation in fluid pipelines. *Control Engineering Practice*, 9:673–682, 2001.
- [15] C. Verde. Real-time detector for multiple leaks in a pipeline. *ISA Monterrey 2002*, pages 57–67, 2002.
- [16] J.E. White and J.L. Speyer. Detection filter design: spectral theory and algorithm. *IEEE Transaction on Automatic Control*, 32:593–603, 1987.
- [17] B.E. Wylie and V.L. Streeter. *Fluid transients*. FEB Press, Ann Arbor, MI, USA, 1983.