Fault tree analysis for data-loss in long-term monitoring networks

J. Dirksen, J. A. E. ten Veldhuis and R. P. S. Schilperoort

ABSTRACT

Prevention of data-loss is an important aspect in the design as well as the operational phase of monitoring networks since data-loss can seriously limit intended information yield. In the literature limited attention has been paid to the origin of unreliable or doubtful data from monitoring networks. Better understanding of causes of data-loss points out effective solutions to increase data yield. This paper introduces FTA as a diagnostic tool to systematically deduce causes of data-loss in long-term monitoring networks in urban drainage systems. In order to illustrate the effectiveness of FTA, a fault tree is developed for a monitoring network and FTA is applied to analyze the data yield of a UV/VIS submersible spectrophotometer. Although some of the causes of data-loss cannot be recovered because the historical database of metadata has been updated infrequently, the example points out that FTA still is a powerful tool to analyze the causes of data-loss and provides useful information on effective data-loss prevention.

Key words | data-loss, FTA, monitoring network, monitoring set-up, UV/VIS submersible spectrophotometer

J. Dirksen
J. A. E. ten Veldhuis
R. P. S. Schilperoort
Department of Water Management,
Faculty of Civil Engineering and Geosciences,
Delft University of Technology,
P.O. Box 5048, NL-2600
GA Delft,
The Netherlands
E-mail: *j.dirksen@tudelft.nl*;
j.a.e.tenveldhuis@tudelft.nl;
r.p.s.schilperoort@tudelft.nl

J. Dirksen
Waternet,
P.O. Box 94370, NL-1090
GJ Amsterdam,
The Netherlands
E-mail: *jojanneke.dirksen@waternet.nl*

INTRODUCTION

Implementation of long-term monitoring networks raises questions concerning data collection, validation, storage, assessment and utilization. UNESCO (2007) gives useful guidelines to those involved in the design or management of monitoring networks in urban water systems. This publication, among others (Mourad & Bertrand-Krajewski 2002; Rieger *et al.* 2005; Ottenhoff *et al.* 2007) stresses the importance of an adequate data validation. Data validation procedures often classify data into three groups: reliable data, doubtful data and unreliable data. In most validation tools 'no data' is considered as a part of the 'unreliable data' group. Since unreliable and doubtful data cannot be used for analysis, data-users are confronted with a dataset full of gaps, potentially limiting information yield. Therefore, prevention of data-loss is an important aspect in the design as well as the operational phase of monitoring networks. In literature limited attention has been paid to the origin of unreliable or doubtful data. Better understanding of the causes of data-loss can provide valuable insight into the (mal-)functioning of monitoring equipment, communication lines, and data storage. As introduced by Schilperoort *et al.* (2008), this insight can eventually be used to maximize data-yield.

In this paper fault tree analysis (FTA) is used as a diagnostic tool to identify all potential causes of data-loss in a monitoring network in an urban drainage system. Although the presented fault tree was developed for a (semi-)permanent monitoring network (monitoring period >1 year) with multiple real-time sensors in an urban drainage network, most steps in the FTA also apply for other types of monitoring networks.

FTA was developed in 1962 in the nuclear industry and has since been applied in numerous industrial areas where extremely safe systems are required (Vesely *et al.* 1981). In literature examples can be found of proactive and reactive approaches of FTA to prevent malfunctioning of

doi: 10.2166/wst.2009.427

systems. The proactive approach focuses on prevention of failure by means of on-line risk calculation using sensors. In these applications fault trees are used to provide computational means for combining logic to analyze system faults. Examples of the proactive approach can be found in:

- Papadopoulos (2003) on aircraft fuel systems and
- Ulerich & Powers (1988) on chemical processes.

In the reactive approach FTA is used as a diagnostic method to examine the cause of failures in order to avoid similar catastrophic failures in the future. Examples of the reactive approach can be found in:

- LeBeau & Wadia-Fascetti (2007) on bridge collapse and
- Nomura (1992) on cooling of a High Activity Liquid Waste tank.

In this paper an application of FTA to monitoring systems is presented which focus on a reactive approach: prevention of data-loss by systematical analysis of causes of data-loss based on historical data.

FAULT TREE MODEL FOR DATA-LOSS IN MONITORING SYSTEMS

Quantitative fault tree analysis is an example of a risk analysis technique that effectively detects potential failure mechanisms and quantifies probabilities of failure of complex systems, such as extensive monitoring networks. The objective of FTA is to identify all possible failure mechanisms that can lead to an undesirable event, the top event of the tree, in a systematic way. In this paper the top event 'data-loss in monitoring systems' is subject of analysis.

There are 4 basic elements in the development of a fault tree: top event, basic events, AND gates and OR gates as illustrated in Figure 1. The choice of the basic events or resolution of a fault tree depends on the level of detail that is required for a specific analysis and availability of data on basic event incidence. A detailed description of the construction of fault trees can be found in NEN-EN-IEC 61025 (2006) and in Vesely *et al.* (2002).

Description of the monitoring network for which a fault tree is developed

Configurations and components of monitoring systems can vary widely, according to monitoring goals, characteristics of the monitored systems, communication lines and organizational context. The fault tree presented here has been developed for analysis of a monitoring network in the wastewater system of the city of Eindhoven, The Netherlands. The network consists of water quantity and water quality sensors such as flow sensors, water level sensors and rain gauges, UVVIS sensors and NH_4^+ sensors. All sensors are on-line sensors, connected to a central data-base by a wireless communication system. A schematic representation of the monitoring set-up can be found in Figure 2. A more elaborate description of the monitoring network is given in Schilperoort *et al.* (2006).

Presentation of main components of data-loss fault tree

Figure 3 shows the fault tree as developed for the described monitoring network. The presented fault tree is only

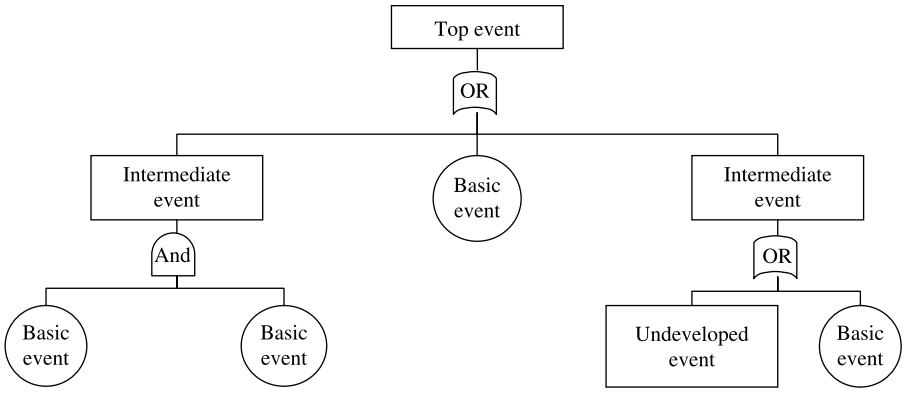

Figure 1 | Elements of a fault tree model.

partially developed in this paper for reasons of clarity; information on deeper levels of the tree can be obtained from the authors. The fault tree has two main branches or failure modes: either no data is present in the database or the collected data is classified as unreliable.

Failure mode: no data

No data implies that no time stamp is registered in the data series where one should have been. This can be caused by a failure of a physical part of the network or by problems in the data communication or the data storage (see Figure 2). In this project, physical components were maintained by the proprietor of the network, the data transfer and the data server were provided by hired third parties. The three events are connected by an OR-gate, because failure of each event individually generates failure at the higher level (i.e. no data), resulting in data-loss. Basic events for 'failure of a physical part of the network' are not shown in Figure 3; examples are: break-down of e.g. a sensor, an antenna etc.

Problems in the transfer of data from a local router to the central router do not necessarily result in data-loss. For data-loss two events need to take place: failure in the wireless data transfer and it is impossible to store the data temporarily at the monitoring location. Because the combination of both events leads to no-data these events are connected by an AND gate. Local storage of data is impossible when there is no local storage facility installed between sensor and router or the installed storage capacity is too small to cover the entire period of network unavailability.

Failure mode: unreliable data

Unreliable data occur when data are delivered from the network, but do not represent the physical process at the monitoring location correctly. An often overlooked cause of unreliable data can be an incorrect data translation. Data translation is necessary to facilitate the transfer of data by means of electronic equipment and wireless data communication. With the use of modules the original monitoring signal is translated from a value with physical meaning to an ampere signal and further to a bit-signal; computer software is used to translate this bit signal back into the original physical value that is ultimately stored in the database. A Site Acceptance Test (SAT) for data translation should confirm the correctness of this translation procedure by comparing database values to local off-line measurements. Every change in the monitoring set-up generally requires a renewed SAT.

The main causes for generation of unreliable data are conditions that prevent a sensor from reproducing the actual physical conditions at the monitoring location. Examples of related basic events are: sensor (location) pollution, sensor drift, sensor maintenance, sensor in wrong position, sensor software problems, sensor damage, etc.

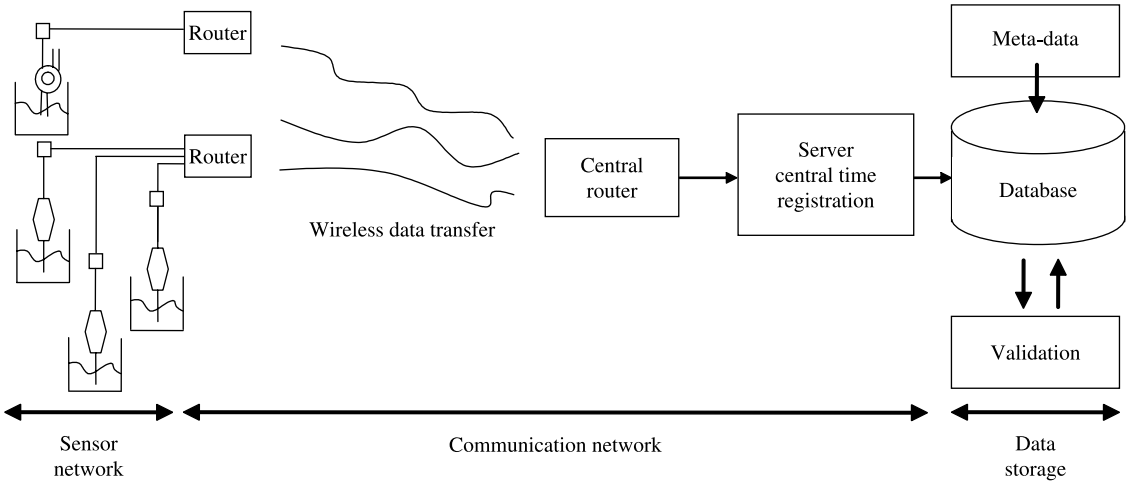

Figure 2 | Monitoring network set-up for the waste water system of Eindhoven city.

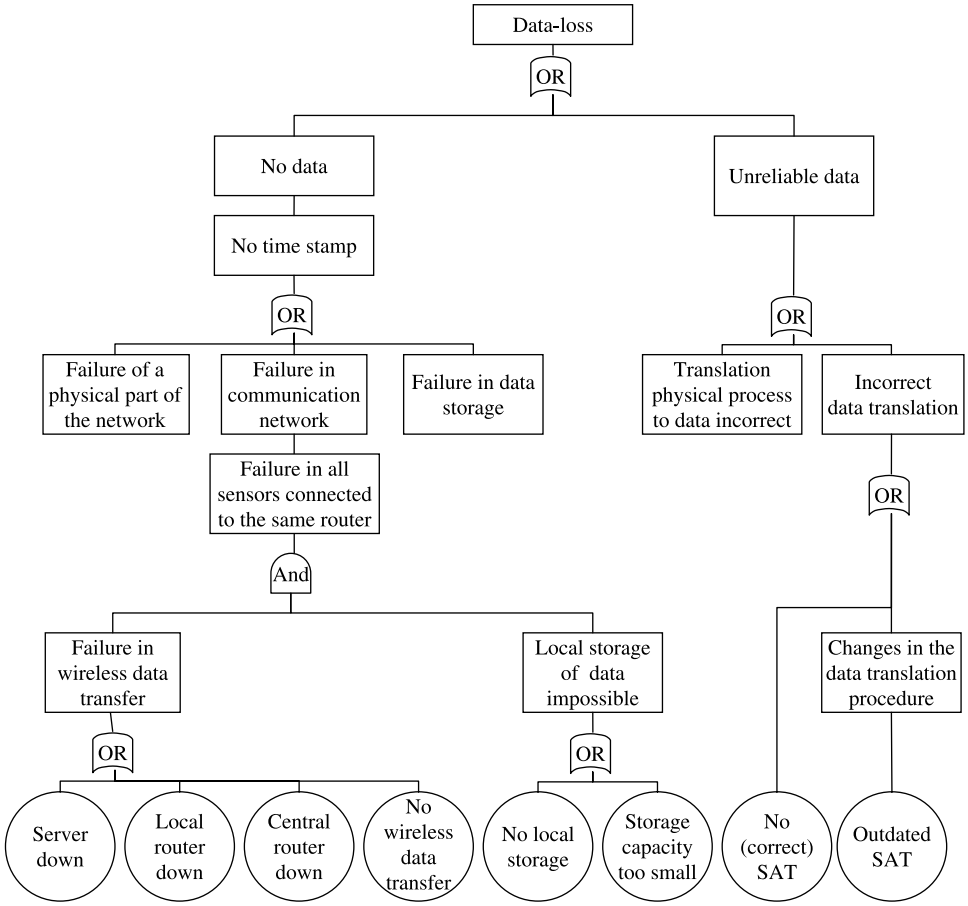

Figure 3 | Data-loss fault tree, partly undeveloped.

APPLICATION OF FAULT TREE ANALYSIS FOR DATA-LOSS

To illustrate the effectiveness of FTA to maximize data yield this paragraph focuses on the analysis of one specific sensor. The goal of the analysis is to identify and quantify the causes of historical data-loss in order to effectively prevent future data-loss.

The considered sensor is a UV/VIS submersible spectrophotometer (Spectro::lyzer, S::can, Vienna, Austria), a water quality sensor that can determine concentration values of e.g. TSS_{eq}, COD_{eq}, $COD_{filtered,eq}$ and NO_{3eq}^{-}, see e.g. Gruber *et al.* (2006). A 1.5-year dataset of TSS_{eq}-data with a monitoring interval of 2 minutes was used for the FTA. A historical metadata file on failures and maintenance activities was available.

Development of a sensor specific fault tree

The general fault tree for data-loss (Figure 3) was used as a starting point for the development of the UV/VIS data-loss fault tree. A list of all evaluated basic events in the fault tree for the TSS_{eq} data-loss can be found in Table 1.

Some basic events were specially added to the fault tree for the application to the UV/VIS data, for instance basic events relating to the cleaning of the lens of the UV/VIS sensor. In order to keep the lens clean, a compressor is added to the sensor set-up to compress air for automatic cleaning. When the compressor breaks down, the sensor is not automatically cleaned every 15 minutes and the data quickly becomes unreliable (basic event: compressor broken down). Besides the automatic cleaning the sensor needs manual cleaning regularly, which results in approximately

Table 1 | Results of the quantitative evaluation of TSS_{eq} data-loss

	OR gate					
Failure mode	**Time span**	**Which sensors**	**Basic event**	**Lost data points (% of total)**	**Duration of lost data (days)**	**# of basic events**
No data	>1 hour	All	Upgrading of the sensor network	1.1	6	1
	>1 hour	All	No wireless data transfer	0.9	5	4
	>1 hour	All	Local router down	1.3	7	1
	>1 hour	One	Failure of a physical part: sensor connection to local router	2.5	14	7
	>1 hour	One	Failure in data storage: data in wrong file	2.3	13	2
	<1 hour	All	No wireless data transfer	0.5	3	>50
Unreliable data			Sensor in maintenance	1.3	7	26
			Compressor broken down	2.9	16	2
			Sensor in wrong position	3.1	17	1
			Unknown cause	11.8	65	12
Total				27.3	150	

3 hours of unreliable data per cleaning operation (basic event: sensor in maintenance).

Since FTA is based on system analysis, the clear representation of interrelations between the elements helped to identify the correct causes of data-loss and even led to the identification of causes of events of which the origin could not be found in the metadata. For instance the fault tree shows that failure of a group of sensors (indicated in Table 1 by 'all') can only be attributed to a shared component, in this case the communication network. A second example is the distinction between data gaps of more than and less than one hour for the failure mode 'no data'. Basic events of the latter are due to wireless data transfer failure: normally local routers send data to the central router once per 15 minutes but if data transfer is impossible (up to four trials), the data of this time-span is lost because there is no local data storage (indicated in Table 1 by '< 1 hour').

Quantitative evaluation of the fault tree

In most quantitative FTA applications the basic events are evaluated based on probability of occurrence. In case of data-loss, probability of occurrence is an improper measure since not only the frequency with which a failure occurs, but also the duration of a failure determines the amount of data-loss. Therefore, the basic events are evaluated based on frequency and duration where duration is defined as a percentage of data that could have been collected but was lost.

In order to determine the frequency and duration of the basic events the following steps need to be taken:

- data validation
- evaluation of the validation results; finding the causes of data-loss using meta data
- calculation of frequency and duration

Because of the measuring set-up and characteristics of the parameter to be measured, TSS_{eq}, it was not possible to validate the data using an automatic tool. Therefore, the data was manually validated based on expert knowledge using rainfall data and metadata on failures and maintenance activities. For safekeeping, doubtful data was considered as unreliable data. Because the same data was used for the validation and evaluation of the data, both analyses were preformed simultaneously.

Interpretation of the results

Table 1 presents data for basic events of the two failure modes in the fault tree. Although the contribution of

individual basic events to the total loss are in the order of a few percent of total potential data yield, the total data-loss amounts to 27.3% or 150 days of lost data. The duration of basic events differs significantly, from less than 1 hour to 17 days. Twelve events could not be attributed to a cause due to a lack of information in the metadata files. These represent almost half of the total lost data points.

For the failure mode 'no data' the main cause of data-loss lies in data transfer problems between sensor and local router. Since the exact malfunctioning component could not be deduced from the data or the metadata, further investigations are needed to identify exact failure and prevent data-loss in the future. Another basic event significantly contributing to data-loss is the storage of data in wrong data files due to inadequate management of the database. A third party was contracted for this service without inclusion of specifications with respect to performance in the contract. Resolving this omission can improve the data yield.

The basic event 'problems in wireless data transfer for a period shorter than one hour' occurred more than 50 times during the 1.5-year period of analysis. Although this only accounts for 0.5 percent of data-loss the consequences for data-users can be problematic since most data analyses require a continuous data-set. The problem can partly be overcome by filling the gaps with synthetic data (Fletcher & Deletić 2007). The basic event 'sensor in maintenance' also has a high frequency of occurrence. Because these data gaps have a time span of roughly 3 hours, filling of the gaps with synthetic data is questionable. Here, sensible planning of manual cleaning of the sensor can increase information yield significantly.

Unlike the previous, the basic events 'compressor broken down', 'sensor in wrong position' and 'local router not in operation' have a low frequency and a long duration. These long durations can be explained by the organizational context of this particular monitoring program. The organization involved was not ready to receive, view and validate the measurement data nor to quickly resolve detected problems. As a result, failures were not identified adequately and correction of failures took longer than necessary. Part of the data-loss could have been saved by a clear project decision structure, priority setting and attribution of responsibilities.

Based on the FTA and interpretation of the results, the following actions are advised in order to prevent data-loss in the future for the presented project:

- Improvements in the organizational context with the aim to view and validate the data regularly and solve detected problems quickly.
- Better updating of meta-data files in order to retrace causes of all data-loss events.
- Sensible planning of manual cleaning operations.
- Detecting the cause of problems in data transfer between sensor and local router. These problems can probably be prevented by the installation of local storage.
- Better management of the database, e.g. by performance-based contracting.
- Improvements in the data communication, e.g. by performance-based contracting.

CONCLUSION AND RECOMMENDATIONS

In this paper FTA is introduced to systematically deduce the causes of data-loss in long-term monitoring networks in urban drainage systems. FTA has shown to be a structured method to detect causes of data-loss because it provides insight into interrelations between system elements through connections in higher levels. This is a clear added value compared to methods that use lists of causes produced by expert knowledge. Application of FTA to data-loss has pointed out how data and metadata can be used to quantify the basic events of data-loss in monitoring networks. Quantitative evaluation of the fault tree indicates the main sources of data-loss and, in combination with the fault tree, can be used to deduce effective solutions for data-loss prevention.

In the presented example FTA is applied in a reactive approach: prevention of data-loss by systematically analyzing the causes of data-loss based on historical data. Even though it became clear that the historical database of metadata was updated infrequently and not all causes of data-loss could be recovered, FTA still proved to be a systematic tool to analyze the causes of data-loss and point out effective solutions for data-loss prevention.

Although not presented in this paper, FTA can also be used proactively when the data can be validated on-line.

On-line validation tools can for example be used to predict failure of the sensor by testing the data on trends. In addition to on-line data validation computer software can also be used to on-line quantify the basic events of the fault tree, allowing effective detection of increases in failure of components in a certain part of the monitoring network. This can eventually help the manager of the monitoring network to detect failures at an early stage.

REFERENCES

Fletcher, T. D. & Deletić, A. 2007 Data requirements for integrated urban water management. The United National Educational, Scientific and Cultural Organization (UNESCO), Paris, France.

Gruber, G., Bertrand-Krajewski, J.-L., De Bénédittis, J., Hochedlinger, M. & Lettl, W. 2006 Practical aspects, experiences and strategies by using UV/VIS sensors for long-term sewer monitoring. *Water Pract. Technol.* **1**(1), doi: 10.2166/wpt.2006.020.

LeBeau, K. H. & Wadia-Fascetti, S. J. 2007 Fault tree analysis of Schoharie Creek Bridge collapse. *J. Perform. Constr. Facil.* **21**(4), 320–326.

Mourad, M. & Bertrand-Krajewski, J.-L. 2002 A method for automatic validation of long time series of data in urban hydrology. *Water Sci. Technol.* **45**(4–5), 263–270.

NEN-EN-IEC 61025 2006 Fault tree analysis (FTA). International standard.

Nomura, Y. 1992 Fault tree analysis of loss of cooling to a HALW storage tank. *J. Nuclear Sci. Technol.* **29**(8), 813–823.

Ottenhoff, E. C., Korving, H. & Clemens F. H. L. R. 2007 Automatic validation of large sets of sewer measurement data. In: *Proceedings of the 3rd International IWA Conference on Automation in Water Quality Monitoring–AutMoNet 2007*, Gent, Belgium, 5–7 September 2007.

Papadopoulos, Y. 2003 Model-based system monitoring and diagnosis of failures using statecharts and fault trees. *Reliab. Eng. Syst. Safety* **81**, 325–341.

Rieger, L., Thomann, M., Gujer, W. & Siegrist, H. 2005 Quantifying the uncertainty of on-line sensors at WWTPs during field operation. *Water Res.* **39**(20), 5162–5174.

Schilperoort, R. P. S., Flamink, C. M. L., Clemens, F. H. L. R. & van der Graaf, J. H. J. M. 2006 Long-term monitoring campaign in the wastewater transport system of WWTP Eindhoven, The Netherlands: the setup. In: *Proceedings of the 2nd International Conference on Sewer Operation and Maintenance*, Vienna, Austria, 26–28 October 2006.

Schilperoort, R. P. S., Dirksen, J. & Clemens, F. H. L. R. 2008 Practical aspects for long-term monitoring campaigns: pitfalls to avoid to maximize data yield. In: *Proceedings of the 11th International Conference on Urban Drainage*, Edinburgh, Scotland, UK, 30 August–6 September, 2008.

Ulerich, N. H. & Powers, G. J. 1988 On-line hazard aversion and fault diagnosis in chemical processes: the digraph + fault-tree method. *IEEE Trans. Reliab.* **37**(2), 171–177.

Vesely, W., Goldberg, F. F., Roberts, N. H. & Haasl, D. F. 1981 *Fault Tree Handbook*. NUREG-0492.US Nuclear Regulatory Commission, Washington, USA.

Vesely, W., Dugan, J., Fragola, J., Minarick, J. & Railsback, J. 2002 *Fault Tree Handbook with Aerospace Applications. Version 1.1*. NASA Headquarters, Washington, USA.