

Monitoring water distribution systems: understanding and managing sensor networks

D. D. Ediriweera and I. W. Marshall

Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK

Received: 5 February 2010 – Published in Drink. Water Eng. Sci. Discuss.: 12 April 2010

Revised: 23 August 2010 – Accepted: 2 September 2010 – Published: 27 September 2010

Abstract. Sensor networks are currently being trialed by the water distribution industry for monitoring complex distribution infrastructure. The paper presents an investigation in to the architecture and performance of a sensor system deployed for monitoring such a distribution network. The study reveals lapses in systems design and management, resulting in a fifth of the data being either missing or erroneous. Findings identify the importance of undertaking in-depth consideration of all aspects of a large sensor system with access to either expertise on every detail, or to reference manuals capable of transferring the knowledge to non-specialists. First steps towards defining a set of such guidelines are presented here, with supporting evidence.

1 Introduction

1.1 Water distribution networks and sensing

Sensor Networks perform an important role in facilitating efficient management of distributed industrial infrastructure. Sensors are used for monitoring operational status of critical assets with the objective of identifying potential issues early on. Benefits of such monitoring include the possibility of performing proactive maintenance leading to significant financial savings, and providing regulated industries with efficient means of maintaining and managing distributed infrastructure within regulatory requirements.

Sensors have been used by the water distribution industry in the past for offline monitoring of assets. Recently however they are being increasingly trialed for online monitoring of water distribution infrastructure (Stoianov et al., 2007). Within this new context sensors are used for collecting information in near real-time for facilitating improved proactive management of pipe systems. Achieving this objective is however highly dependent upon the operational performance of sensor systems in discussion.

Proper design, deployment, and management of sensor systems are therefore critical for achieving the level of efficient and reliable operation necessary. Acquiring skills and expertise to achieve this can sometimes be expensive and difficult, leading to compromised system design and operation.

Non-optimal design choices not only lead to poor performance of the sensor system itself, but also compromise any system which relies on its information.

Deploying sensor networks for monitoring distributed infrastructure invariably introduces a second complex system into the picture; this being the sensor network itself. Being a distributed system; a sensor network will often carry management complexities similar to that of the system which it is designed to monitor, and as a result, to realise efficient and sustainable performance, it will require well designed tools for supporting constant system monitoring (Tierney et al., 2001).

1.2 Outline of work

A large-scale industrial sensor network deployed for online monitoring of pressure and flow measurements of a wide area water distribution and supply system is investigated. System performance, system design, system deployment, and system management are examined. The work aims to identify design and operational limitations which the system encountered during its initial 3+ year deployment period and attempts to identify which of such limitations were related to poor design-time choices.

A series of correlation tests between system failures and possible contributing factors are presented. Potential failure mechanisms are discussed, and reasons behind such failures analysed. Findings and observations are presented as a set of design guidelines aimed at helping non-specialist engineers make high quality design choices in the future. An insight



Correspondence to: D. D. Ediriweera
(damjee@gmail.com)

into several non-obvious design pitfalls observed within the investigation are outlined to help future designers avoid the same.

2 Related work

Work published by (Stoianov et al., 2007) provides a useful insight into the emerging role of sensing within water distribution systems. Work presented discusses a solution for monitoring large diameter bulk-water transmission pipelines through wireless sensor networks for burst and leakage detection. Increased spatial and temporal resolution of hydraulic and acoustic/vibration data is claimed to provide significant proactive and reactive maintenance capability allowing for potentially large financial savings. Field results indicate key observations including the necessity for robust sensing hardware, decoupling of data collection from communication, and improved time synchronisation.

A different application of sensing with water distribution infrastructure is presented by (Qian et al., 2007). The paper highlights the importance of security in urban water distribution infrastructure, and discusses a sensor solution for improving its safety and protection via an optimisation between off-line and online monitoring. Other work such as environmental sensing for monitoring drinking water quality (Ailamaki et al., 2003), and project SmartCoast for monitoring water quality in fresh water catchments (B O'Flynn et al., 2007) do also exist which provide further evidence of the growing application of sensor networks within the drinking water industry.

A specialized communications model for sensor based monitoring of water distribution networks is presented by (Lin et al., 2008). The work specifically addressed challenges encountered in underground to aboveground radio propagation when sensors are placed within concrete and cast iron chambers; a typical placement of sensor nodes common within many water distribution networks. The work discusses different antenna placements, antenna sizes, frequencies and strategies for establishing the necessary uplinks, and provides useful insights into practical communication challenges which need consideration in real world sensor deployments.

Common sensor network architecture for equipment failure prediction through vibration monitoring is discussed by (Krishnamurthy et al., 2005). The architecture is trialed within two industrial surroundings: a semiconductor manufacturing plant; and a North Sea oil tanker. While the work may not directly relate to water distribution, findings are relevant as they are presented in terms of general design insights. Important findings include relationships observed between different sensor hardware configurations and their power efficiency. It is claimed that hardware with greater RAM and I/O bandwidth were more power efficient as they required less software intelligence for resource management.

Based on the literature it becomes evident that sensing within water distribution systems is useful for many applications ranging from burst and leakage detection, infrastructure security, water quality monitoring etc. However at the same time only few guidelines appear to be available on "how" such sensing should be achieved. Available models often appear to address design time issues and rarely discuss illusive pitfalls which often become apparent at deployment and operation time. Guidelines presented here are therefore aimed at assisting designers avoid issues which are illusive and subtle at design time, but impactful during live operation. These guidelines are derived by analyzing the performance of a live sensor network in retrospect, and hence should effectively capture runtime limitations and their relationship to design time choices.

3 Survey details

3.1 Information used

A multi-faceted analysis has been undertaken where a variety of variables and potential contributing factors have been examined to understand "how" some may have affected system performance. The main analysis has been performed based on the original dataset collected by the sensor network over the studied period. Following additional information and knowledge has also been made use of in the analysis.

- Discussions with the sensor network operator, and analysis of the water distribution system.
- Analysis of the sensor locations and their distance to communication towers.
- Analysis of local weather conditions.
- Experience from other sensor networks including monitoring systems in telecommunications.

3.2 System configuration

The examined system is a wide area sensor network deployed for monitoring of a water distribution and supply system. The sensor net consists of 520 plus sensor nodes monitoring flow and pressure of a pipe system. These nodes are distributed within roughly 50 km². The majority of the sensor nodes are deployed underground within manholes located on public roads. All nodes including their sensors, radios and loggers are commercially manufactured and are designed to be water proof. Electronics within these were custom made with the possibility for minor imperfections. Each sensor node is equipped with a GSM modem capable of GPRS data connectivity. Data collected by individual sensor nodes are relayed via a public GSM/GPRS network. Sensor nodes are powered using a battery pack with an estimated lifetime of approximately twenty four months.

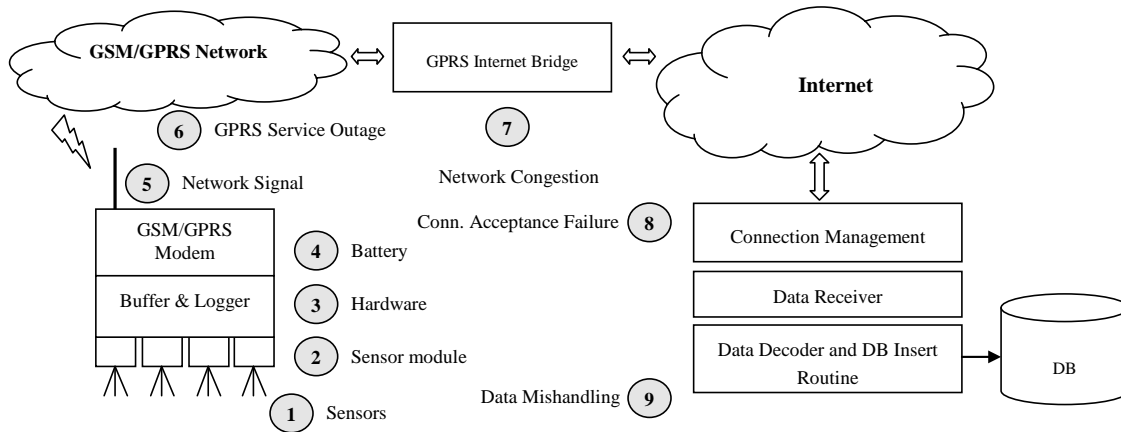


Figure 1. The diagram outlines the configuration of the sensor network being investigated. An abstracted view of the process is captured to maintain clarity. Circled numbers indicate possible failure points within the infrastructure which are relevant to the analysis.

Table 1. Sensor net performance summarised by year.

Year	Total Days Operational	Total Failings	Loggers Operational	Loggers Failed	Logger Failure % (LFP)	Missing Data % (MDP)	Total Records Lost	Records Lost per Logger per Day (RL/L/D)
2006	364	822	479	321	0.670	8.03	1 196 956	6.86
2007	364	1242	462	401	0.867	7.45	1 389 133	8.26
2008	365	2527	430	349	0.811	32.8	4 465 913	28.45
2009	71	137	387	289	0.746	0.32	21 987	0.80

Each sensor node is designed to support up to a maximum of 4 data channels, with each channel supporting a single sensor. 68% of the nodes deployed were using a single connected channel while the remaining 32% were using both. Sensor nodes with two active channels were monitoring both flow and pressure, while nodes with a single active channel were measuring either flow or pressure.

Data path configuration of the system is highlighted in Fig. 1. Sensor nodes are configured to record readings every 15 min. Pressure readings are recorded as snapshots while flow readings are averaged over several measurements within a 15 min period. The loggers report back to the central data collector every 30 min. The dataset used for the analysis includes in excess of 76 million records over the 3+ year period from 2006 to 2009. 49 million of these readings were of pressure, while 27 million were flow. The data has been recorded from 529 distinct sensor nodes with 697 active channels.

4 The analysis

4.1 Missing data overview

Table 1 provides a summary of yearly performance of the system. All failures in the table refer to events of missing data within the dataset. A missing event in this instance is

defined as a gap of greater than 15 min between two consecutive measurements recorded for a specific channel. This is a conservative measurement as it does not take in to consideration any records which continued to be missing at the end of the studied period. Key indicators to be noted from Table 1 are: Logger Failure Percentage (LFP); Missing Data Percentage (MDP); and Records Lost per Logger per Day (RL/L/D).

LFP identifies the percentage of loggers which at least failed once during a given year. Over the studied period LFP appears to be fairly stable indicating that overall conditions such as manufacturing quality of the hardware, and long term environmental conditions were generally stationary during the considered period. MDP identifies the volume of data missing in a given year as a percentage of the total data which should have been recorded during the same period. MDP is therefore an important indicator of overall system performance. MDP for year 2006 and 2007 are stable at around 6%–8%, but dramatically increases to 32% in 2008. This indicates a much higher level of failures during this year. MDP for 2009 indicates a big improvement at 0.8%, but may not be indicative as the figures are only for a short period of the year. RL/L/D identifies on average the number of records missing per logger per operational day. RL/L/D appears to be consistent with MDP. It peaks in 2008 and then falls off in 2009. In summary these findings indicate system

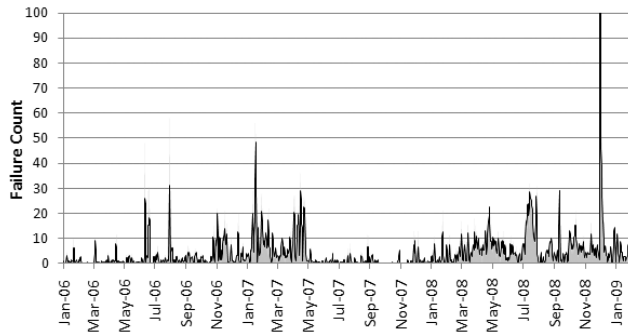


Figure 2. Short-term (<6 h) failures vs. date.

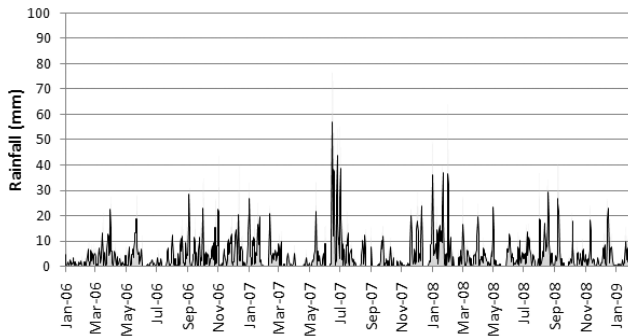


Figure 3. Rainfall vs. date.

performance to be generally poor incurring significant loss of data during each year excluding 2009 (where only a short period of the year is analysed). 2008 records an unusually high loss rate at 32%.

4.2 Missing data by date

Figure 2 illustrates the distribution of the short term (less than 6 h) missing data events over the 3+ years analysed. The plot indicates short-term failures to be fairly clustered with some periods illustrating higher numbers of failure than others. Peak periods can be identified as Jan-07 to May-07, Mar-08 to Aug-08, and Oct-08 to Dec-08. The calmer periods with fewer failures occur between Jan-06 to Nov-06, and May-07 to Jan-08. No lull periods appear to have occurred Jan-08 onwards.

Figure 3 plots rainfall recorded during the period from Jan-06 to Jan-09. This has been calculated as the average from 6 rain gauges in the general area. Rainfall is of interest due to the effects which it has on radio wave propagation and the possibility of it causing flooding of underground sensor nodes. Upon comparing Figs. 2 and 3, none of the peak rainfall events appear to significantly correlate with failure events over the 3+ year period. A large rainfall event in Jul-07 which is recorded to have caused widespread flooding in the area does not appear to have triggered significant sensor fail-

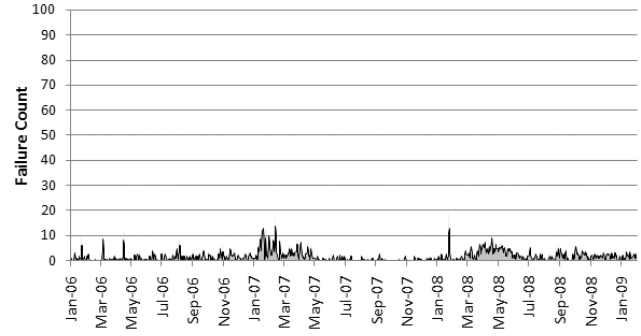


Figure 4. Long-term (>6 h) failures vs. date.

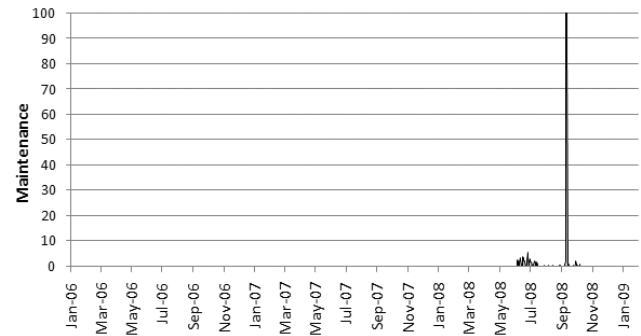


Figure 5. Sensor network maintenance vs. date.

ures either. Based on this evidence it is therefore appropriate to exclude rainfall as a major contributor of sensor failures.

Figure 5 plots maintenance work carried out on the sensor network between Jan-06 to Jan-09. This is of interest as maintenance may cause system disruptions. It is in fact found that short term failures during Jul-08 and in Sep-08 closely correlate with such maintenance. This strongly backs the possibility of these failures being caused as the result of engineers performing maintenance on sensor devices.

Figure 4 plots long-term errors lasting for greater than 6 h. Two clusters of long-term errors appear from Jan-07 to May-07 and from Mar-08 to Jul-08. Based on operator feedback these periods were found to correlate with battery replacement schedule and could therefore indicate devices switching off due to battery depletion. Specific battery replacement records were not available for verifying this. Long-term errors which do not fall within these periods were potentially caused due to out-of-synch battery depletion or sensor hardware failures.

4.3 Missing data by duration and event size

Figure 6 plots failure events grouped by event duration and event size. Two distinct distributions are visible: short-term errors "<24 h", and long-term errors ">24 h". Long-term errors peak at ">30 days" and tail off at "2 to 3 days". It is interesting to note that 99.5% of these long-term errors affect

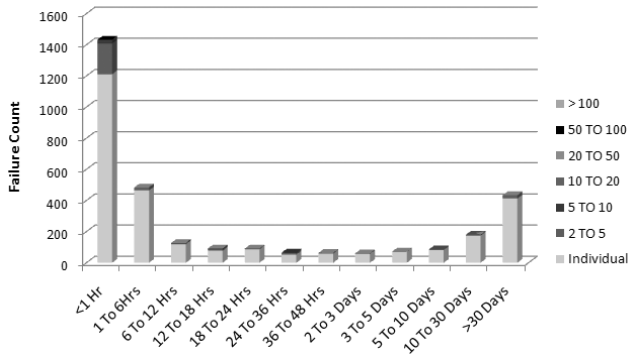


Figure 6. Missing data events by duration and event size.

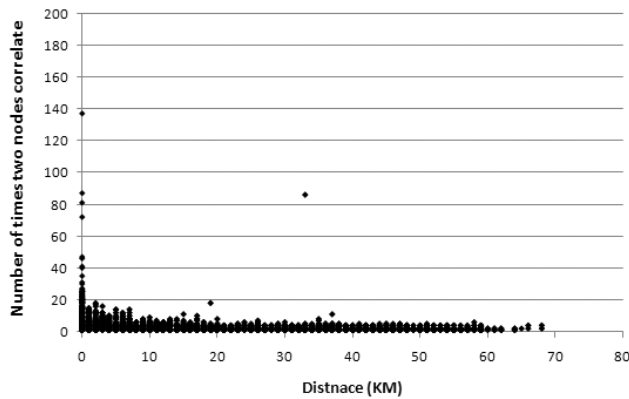


Figure 7. The relationship between the number of times two sensor nodes failed together versus the distance between them. A one hour correlation window is used.

individual sensor nodes alone, which supports the earlier observation that such failures are likely to be caused due to sensor hardware or battery failures.

Considering short-term errors which peak at “<1 h” and tail off towards “24 to 36 h”, only 84% of them are affecting individual loggers. These 84% are likely caused by a host of issues such as sensor node maintenance, temporary signal-losses, and ad-hoc occurrences such vehicular traffic, or radio interference. It is difficult to determine specific causes as no error logs from the sensor nodes were available. Considering the remaining 16% of short-term errors, 13% has simultaneously affected groups of 2 to 5 loggers, and the remaining 3% affecting larger groups of 5 to 100 loggers. Mechanisms of failure here are different from the individual logger failures discussed above. Short-term failures affecting small groups are most likely caused due to temporary conditions affecting small localities, e.g. local GSM signal conditions. This is supported by the illustration in Figure 7 which plots the relationship between the number of times two sensor nodes failed together, versus the distance between them. Based on Fig. 7 it is evident that strongly correlating sensor nodes are almost always located closer to each other indicating that

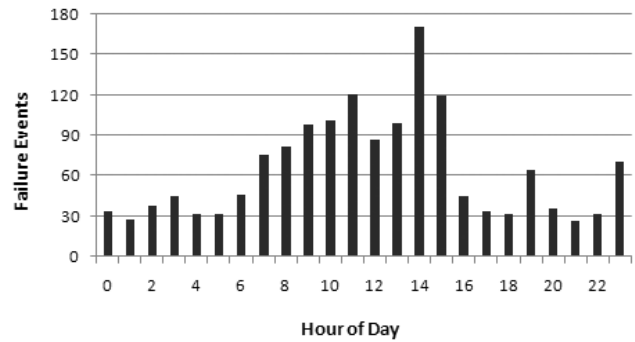


Figure 8. Weekday missing data events against hour of day.

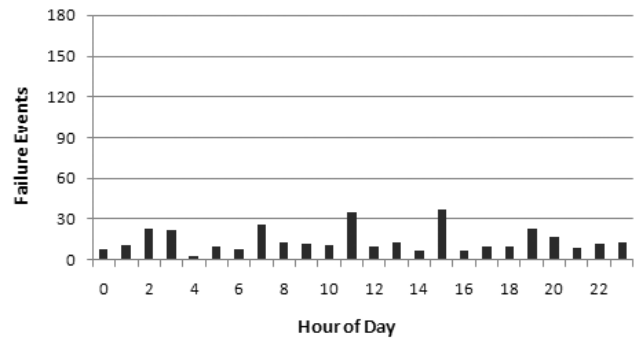


Figure 9. Weekend missing data events against hour of day.

these short-term group failures are locality related. On the other hand short-term failures affecting larger groups such as 50 to 100 sensor nodes are caused due to network wide events such as GPRS issues, back-end failure, or backend data mismanagement issues.

4.4 Missing data by hour of day

Figures 8 and 9 plot missing data events against the time of day during which they occurred. Figure 8 plots failures during weekdays, while Fig. 9 plots the same for weekends. Two distinct patterns have emerged. During weekdays a strong correlation is seen between failure events and common working hours. In relation to the sensor network, this could indicate an association of failures with GSM/GPRS network traffic which would be expected to peaking during similar hours. While no traffic data from GSM network operators were available for analysis, the hypothesis gathers pace due to the fact that the correlation disappears during weekends where no peak traffic would be expected during similar hours.

Alternatively, the cause for the distinct distribution in Fig. 8 is also explainable by an entirely different phenomenon. Considering that common business hours and GSM/GPRS peak hours also closely correlates with typical working hours of network engineers managing the sensor

Table 2. Overall breakdown of failures by estimated type.

Type	# Failures	# Lost Records	% Failures	% Lost Records
Battery (LT)	455	839 676	9.62%	11.87%
Hardware (LT)	1170	6 051 568	24.74%	85.54%
Network (ST)	538	32 250	11.38%	0.45%
Maintenance (ST)	659	37 980	13.94%	0.54%
Unclassified (ST)	1906	112 515	40.31%	1.59%

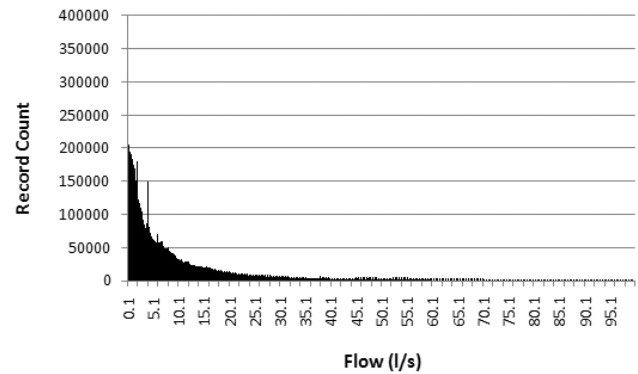
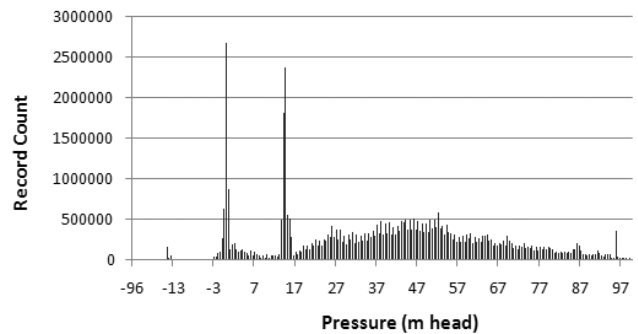
(LT) – Long-Term, (ST) – Short-Term

network, it is plausible that these failures are caused due to maintenance work which may not have been recorded. While this is a possibility, it is an unlikely one since short-term failures largely appear to be distributed throughout the period of study as illustrated by Fig. 2. The original hypothesis relating to GSM/GPRS network traffic therefore looks more likely, but it is yet to be verified as no GSM/GPRS network traffic information or error logs from loggers in question were available for further analysis.

Table 2 provides a classification of failures by category. These are only estimates given that they are based entirely on the analysis of data. These estimates cannot be confirmed since no error logs were available from the sensor nodes. Table 2 nevertheless provides a rough idea of the causes of the failures within the studied system. While over 65% of the failures are short-term, they appear to only cause just above 2.5% of the data loss. A large percentage of the missing data are therefore due to long-term failures. This however does not necessarily identify which type of events are most damaging as even a short-term failure during a critical time, at a critical point within the water distribution infrastructure can be more damaging than a long-term failure. Unclassified failures are partly due to maintenance work not being recorded, and the remaining possibly caused by mishandling of data at the backend.

4.5 Data quality analysis

Figure 10 plots the distribution of flow values within the dataset. The distribution appears free of obvious errors with a generally smooth tail and a clear peak at value zero. Zero flow is an acceptable reading as this occurs on certain pipes during off-peak hours. However, it is found that approximately 14% of these flow zeros measurements are erroneous – this is based on the assessment that they occur continuously for periods of over 24 h. Figure 11 plots the distribution of pressure values within the database. Three peaks are visible; the first and the highest occur at zero. Zero pressure is however an invalid reading under normal operational conditions of a pipe system and is therefore considered as an erroneous reading. The second peak occurring at 14 is also unexpected, but given the magnitude of the peak it is clear

**Figure 10.** Distribution of flow values.**Figure 11.** Distribution of pressure values.

that it is unlikely to be genuine. The distribution also indicates some pressure readings to be negative which are also non-feasible. In summary, it is apparent that the general quality of pressure readings is of poor standard whereas flow readings appear to perform significantly better.

4.6 Overview of the data analysis

The overall analysis reveals 12% of the data which the system should have collected to be missing.

A further 7% of the data are zero valued with no explanation of being genuine, and are therefore considered erroneous. 2% of the data are negative values, and 1% impossibly large values. In total this amounts to a figure high as 22% of the data intended to be collected, either to be missing, or erroneous.

5 Rules for future design

A fundamental oversight in the design of the current sensor system is its lack of support for effective monitoring and management. The sensor system being a distributed entity requires constant management. Necessary tools and mechanism for achieving this should therefore be designed in to the system. Within the current system it is clear that designers were unaware of the complexities of managing a large sensor network, and as a result failed to provide even the most

fundamental management information such as the error logs. Resulting from this lack of information the operators of the sensor network were unaware of the scale of the performance lapses within the system. To overcome such limitations it is vital that system designers envisage how a system can be managed when it is deployed, and use this insight to integrate necessary tools to support such management. At a minimum such support should include the possibility of monitoring sensor nodes continuously and red-flagging possible issues as they occur. Error logs should be available for further investigation of any issues. Facilities should also be introduced within sensor management systems for automated monitoring of battery performance and scheduling of battery replacements. This would help minimise downtime of sensor nodes due to battery failure. It was surprising to discover that batteries within the current system were hardwired to sensor nodes, making their replacement a complex process. Issues such as hardwired batteries and unavailability of sensor error logs suggest significant lapses in requirement specifications by sensor network operators to the equipment vendors.

The selection of a suitable communication medium and appropriate communication protocols are other aspect of sensor network design which requires significant consideration. Current system operators have opted for GSM/GPRS solution while the industry norm has largely remained focused towards GSM/SMS. Although such forward thinking is valuable, extreme caution is necessary to fully understand their implications. As an example, the effects of peak GSM/GPRS network traffic discussed earlier are potentially a hidden phenomenon with considerable adverse side-effects. The existence of such an elusive problem suggests the importance of undertaking performance evaluations such traffic profiling, and quality of service checks on any third party infrastructure which is planned to be used.

The selection of a proper communication strategy can itself be as vital as selecting the communication medium itself. Within the current system all sensor nodes are polled every 30 min starting from the top of the hour. This approach of simultaneously polling all sensor nodes may itself be sufficient to cause low-capacity GSM/GPRS cells to be congested, creating traffic errors, and connection issues. More importantly, this approach is most likely to create bottle necks at the data collector which would need to handle incoming connection requests from all sensor nodes simultaneously, leading to potential data mishandling and other communication errors within the backend system. The simplest approach to overcome this would be to stagger the polling times such that sensor nodes report back at different individual times during each 30 min block. To compliment such an approach it is also useful to avoid the use of standard times for reporting back data, especially when third party infrastructure is used. The rationale being, the probability of standard times such as the top of the hour, fifteen past, or midnight being used by other systems on same network is much higher compared to non-standard times such as for example three minutes past

the hour. Using such non-standard times increases the probability of avoiding traffic being generated by other systems.

6 Conclusions

The study has revealed up to a fifth of the data collected by the evaluated sensor system to be either missing or erroneous. Contributing factors included lack of expertise in system design leading to poor design choices, lack of management support built-in to the system, weaknesses in the specification given to the equipment vendor by the system owners, back end IT system failures, and failing transducer hardware. Findings in retrospect reveal important aspects which require considerable attention when deploying large sensor systems. These include the recruitment of necessary expertise, design of integrated management tools, careful evaluation and selection of communication mediums, identification of suitable communication strategies, and undertaking of management as a planned activity. Future work on developing these guidelines to formulate a more comprehensive rule set is currently planned.

Acknowledgements. This work is/has been funded by project Neptune (EPSRC Reference: EP/E003192/1).

Edited by: J. Vreeburg

References

- Ailamaki, A., Faloutsos, C., Fischbeck, P., Small, M., and VanBrienen, J.: An environmental sensor network to determine drinking water quality and security, *SPECIAL ISSUE: On sensor network technology and sensor data management*, ACM SIGMOD Record, 32, 4, 47–52, 2003.
- Krishnamurthy, L., Adler, R., Buonadonna, P., Chhabra, J., Flanagan, M., Kushalnagar, N., Nachman, L., and Yarvis, M.: Design and deployment of industrial sensor networks: experiences from a semiconductor plant and the north sea, *Proceedings of the 3rd international conference on Embedded networked sensor systems*, 64–75, 2005.
- Lin, M., Wu, Y., and Wassell, I.: Wireless sensor network: Water distribution monitoring system, *Radio and Wireless Symposium*, 2008 IEEE, 775–778, 2008.
- O'Flynn, B., Martinez, R., Cleary, J., Slater, C., Regan, F., Diamond, D., and Murphy, H.: 32nd IEEE Conference on Local Computer Networks 2007, 815–816, 2007.
- Qian, Z., Gao, J., Wu, W., Hou, X.Q., Wu, Y., Tao, L., and Ran, R.: Harbin Water Supply Network Monitoring and Evaluation, *IEEE International Conference on Networking, Sensing and Control 2007*, 345–349, 2007.
- Stoianov, I., Nachman, L., Madden, S., and Tokmouline, T.: PIPENET a wireless sensor network for pipeline monitoring, *Proceedings of the 6th international conference on Information processing in sensor networks*, 264–273, 2007.
- Tierney, B., Crowley, B., Gunter, D., Lee, J., and Thompson, M.: A Monitoring Sensor Management System for Grid Environments, *Cluster Comput.*, 4, 1, 19–28, 2001.