Towards a Novel Early Warning Service for State Agencies: A Feasibility Study

Ioannis M. Dokas[1], Richard J. Wallace, Radu Marinescu, Syed Imran, Franclin Foping

Cork Constraint Computation Centre, University College Cork, Cork, Ireland

Abstract

The paper presents the results of a feasibility study in need to determine if viable, effective and reliable web based early warning service in the domain of drinking water utilities can be developed. The early warning service is intended to be used by many actors in the Republic of Ireland including water utilities managers, the Irish Environmental Protection Agency personnel and the Irish Health and Safety Executive personnel. The results of this feasibility study include an assessment of whether or not a statement of requirements for a novel early warning service can be realised using a set of existing technologies. Based on these results it is concluded that it is feasible to develop a reliable and effective web-based early warning service for drinking water utilities by adapting technologies derived from multiple disciplines including safety engineering, artificial intelligence, software and web engineering.

[1] Corresponding Author: E-mail: i.dokas@4c.ucc.ie

1. Introduction

The contamination of drinking water has been identified as a major problem in many countries. Communities large and small, including those in some of the world's most developed countries, have been affected by contaminated drinking water [8]. In Europe, for instance, 7,960 confirmed cases of cryptosporidiosis were reported during 2005. Ireland with 13.75 and the UK with 9.26 per 100,000 cases have reported the highest incidence rates [18]. Recently, the city of Galway in Ireland experienced the adverse effects of having the drinking water contaminated due to cryptosporidium parasite caused by human sewage [3] resulting to at least 180 serious cases of cryptosporidiosis.

Some waterborne disease outbreaks are related to failures in water treatment and distribution systems. A characteristic example is the 1993 cryptosporidium outbreak in Milwaukee, Wisconsin in the USA caused by cryptosporidium oocysts that passed through the filtration system of one of the city's water-treatment plants [15]. It was estimated that over the span of approximately two weeks 403,000 residents in the Milwaukee area became ill with stomach cramps, fever, diarrhoea and dehydration caused by the pathogen and 69 deaths were attributed to this outbreak [5].

Early warning services for water utilities, a domain that also falls under the category of critical infrastructures, can been seen as a very useful tool in supporting proactive risk management strategies by enhancing existing feedback control mechanism for maintaining water quality and by supporting the dissemination of important information to the stakeholders in a timely manner. However, a characteristic of the drinking water contamination problem is that it is often detected only after a health crisis, when people have fallen ill or died as a result of drinking unsafe water [8]. This implies that more effective and reliable proactive strategies are needed, which must be supported by more effective early warning services.

1.1 Early Warning Systems

Early warning services are delivered to appropriate stakeholders by early warning systems in a wide spectrum of domains. These domains can be considered to be members of a wider taxonomy of systems like for example environmental, industrial, natural, sociotechnical, military, financial and medical systems. Important concepts for early warning systems such as “hazard", “risk", “threat", “safety", “vulnerability", “warning signal" are defined differently due to the different nature and characteristics of each domain. The variability of the characteristics in different domains is one reason for the non-existence of a universally acceptable definition of what exactly early warning systems are [6].

United Nations has recommended that in order to be effective, early warning systems have to integrate four elements (i) a knowledge of the risks faced; (ii) a technical monitoring and warning service; (iii) the dissemination of meaningful warnings to those at risk; and (iv) public awareness and preparedness to act. Failure in any one of these elements can mean failure of the whole early warning system [12].

In the domain of drinking water supply a description of the goal of early warning systems is given in [9; 11]. These authors defines early warning systems as an integrated system for monitoring, analyzing, interpreting, and communicating monitoring data, which can then be used to make decisions that are protective of public health and minimize unnecessary concern and inconvenience to the public.

1.2 Previous Work, Aims and Goal

A review of the literature has identified only one representative example of an early warning system in the domain of water treatment and supply. The WaterSentinel system [20] is the result of a program developed by the U.S. Environmental Protection Agency for designing, deploying, and evaluating a model contamination warning system for timely detection and appropriate response to drinking water contamination threats and incidents that would appear to have broad application to drinking water utilities. In this work the term "contamination warning system" is used instead of the early warning system term reflecting the recognition that a reliable system providing early warning of a contaminant prior to human exposure with public health impacts using today's technologies may not be possible.

In this paper we argue that it is feasible to develop a novel type of early warning service. The main goal of this early warning service is to assess the possibility, or the likelihood, of occurrence of operational problems and failures in a timely manner and through that to assess and inform the stakeholders about the risk of a drinking water contamination incident.

The novel characteristic of our approach in comparison to previous works is that we adapt a system theoretic approach to early warning services that takes into consideration the need to monitor not only critical operational parameters in the water utility but also critical parameters of the components of their wider sociotechnical system in which it is empeded. Ideally, the proposed novel early warning service will make the boundaries of safety performance of water utilities visible to the stakeholders and decision makers and will help them implement, in a timely manner, a strategy to reduce the forces that drive water utilities beyond their safety boundaries.

Later in this paper, the results of a qualitative feasibility study are briefly presented to determine if viable, effective and reliable web based early warning service providing early risk identification in water utilities can be developed. The proposed early warning service is currently in the design phase and is intended to be developed and to be used in the Republic of Ireland by many actors in all levels of the wider sociotechnical system of water utilities. Intended users are the managers of water utilities, Irish Environmental Protection Agency personnel and Irish Health and Safety Executive personnel. The results of this feasibility study include an assessment of whether or not the technology needed for developing such a service exists, given an outline of early warning service requirements.

The paper continues with a description of the requirement analysis followed by a brief description of the stakeholders' needs and of the functional requirements of the services. A description of a set of selected technologies is given in Section 3, based on which the development of the early warning service will be made. Section 4 provides a discussion of the results as well as concluding remarks.

2. Requirements Analysis

The requirements analysis for the early warning service was conducted following a three phased approach.

1. Familiarisation.
2. Analysis and Elicitation.
3. Documentation.

Each phase of the requirements analysis is composed by several steps. A brief description of each step is available in Table 1.

2.1 Stakeholders' Needs and Functional Requirements

Stakeholders' needs describe what stakeholders require from the early warning service. They therefore dictate what type of services the early warning system must provide. The needs were identified during the analysis phase through face to face meetings.

The needs are shown in Table 2. In Figure 1, a UML Use Case diagram is shown to depict the functional requirements as desired features of the early warning system. This diagram provides a model of the interactions between the early warning system and the entities which will interact

Table 1. Phases of requirement analysis

Phase	Steps	Aim/Description
Familiarisation	1	Conduct literature review of accident reports on drinking water disasters. The aim is to get familiar with the water treatment domain, and comprehend the complex combination of circumstances that may lead to waterborne diseases outbreaks and to drinking water contamination disasters.
	2	Understand the characteristics of water utilities operations based on a systems theoretic approach; realise and understand the different hierarchy levels of the organization that water utilities belong to.
	3	Identify and apply existing systems theoretic risk and accident analysis approaches. Understand accident causation and the inherent complexity. Define the actors affecting the "safety health" of water utilities.
Analysis and Elicitation	1	Develop a conceptual model of the sociotechnical system for the case of water utilities in Ireland. Identify key actors of the early warning service.
	2	Elicit the requirements of actors through face to face meetings
	3	Comprehend the degree to which assets like sensors and monitor systems exist and are in use at the water utilities.
Documentation	1	Document the needs of stakeholders and the features of the early warning service.
	2	Relate the needs of stakeholders with the features of the early warning service.
	3	Document the functional and non functional requirements.

Table 2. Stakeholders' Needs

No	Need
1	To be accessible by all stakeholders which are scattered in Ireland
2	To provide important and relevant information to each stakeholder
3	To allow the profiling of each and every water utility and monitoring of critical operation parameters
4	To allow the modelling of knowledge about causes – failures –effects
5	To allow updating of knowledge related to causes – failures –effects
6	To allow estimation of the potential of failure of water utilities, given some evidence and to reason potential of failure with uncertain data
7	To notify the stakeholders when the estimation of a potential failure is considered to be "high"
8	To allow as inputs real time sensor data as well as to allow updates of the status of the components of the water utilities by water utilities personnel
9	To provide advice and emergency response procedures to the stakeholders

Fig. 1. UML Use Case Diagram depicting functional requirements as desired features of the early warning system

with it and is a part of the documentation phase of the requirements analysis.

3. Selected Technologies for Early Warning Service

For the feasibility study a set of technologies were selected and were further studied to determine if they could be used to develop the novel early warning service. These are described in the following sections.

3.1 Risk Analysis Techniques and Domain Specific Modeling

The effectiveness of an early warning system in preventing facility or system failures depends on modeling and representation of authentic and adequate knowledge of involved risks. Therefore, in order to prevent accidental failure effectively, an early warning system must embody knowledge of hazards and vulnerabilities. Thus the need for knowledge modeling and representation is very much of concern to anyone evolved in delivering early warning services whose goal is to prevent system failures.

It can be argued that the problem of knowledge-intensive modeling of how a system or a facility can fail is addressed by common risk analysis techniques. For example, techniques such as, Fault Tree Analysis, Failure Modes and Effects Analysis, and Cause-Consequence Analysis, that were originally developed for reliability and risk analysis, can also be used to prevent system or facility failure [16].

Fault tree analysis is an analytical technique in which an undesired state of the system is specified and the system is then analyzed in the context of its environment and operation to find all realistic ways in which the undesired event can occur [19]. Fault trees can be used to analyze a single unwanted event like a failure or an accident and to determine all combinations of events that will result in it. Each unwanted event is decomposed into the events that must occur to produce that effect. Eventually, the decomposition will result in component failure modes where the analysis usually ends. The unwanted event can be quantified, if required, in terms of a frequency or probability of occurrence, given that the frequencies of the component failure modes are known.

Failure Modes and Effects Analysis is a systematic procedure by which each potential failure mode in a system is analyzed to determine its effect on the system and to classify it according to its severity [1]. It provides tabulation of equipments, components and their associated single point failure modes, consequences, safeguards, and related actions listed.

Cause-Consequence Analysis provides a model containing knowledge about the causal aspects as well as about their consequence aspects of failures. Each knowledge base contains heuristic as well as systematic knowledge about the temporal, probabilistic, and causal aspects of various faults that can occur in a given system [16].

When developing an early warning system there is a need for the representation of knowledge models expressing accident causation at the system level, meaning that there is a need to generate computer code from the models. If the problem domain of an early warning system is “small”, its functionality and its reasoning can be expressed directly in computer code without creating models. However, when the problem domain is “large” and complex there is a need to minimize the gap between the necessary knowledge modeling and code generation.

One solution to this problem is to use Domain-Specific Modeling. Domain-Specific Modeling provides a higher level of abstraction than "typical" Model Driven Development approaches in order to prevent errors in earlier stages of development by not allowing the illegal designs or designs that don not follow the underlying architectural rules, thus producing the code free from careless mistakes, syntax and logical errors [14] without any manual code editing. Domain-specific models can also be used to generate other models and to configure other systems perhaps in combination with other enabling technologies by providing the domain-specific Application Process Interfaces. In so doing they can enhance productivity and generating solutions automatically.

Thus, it can be concluded that knowledge of how systems or facilities can fail can be elicit and modeled using widely known reliability analysis techniques. Further, Domain-Specific Modeling gives us the ability to generate computer code directly from these models and thus to form the final product at the system level (i.e. the early warning system). This can make early warning systems development more accessible to people that are not familiar with the implementation technology.

In short, it can be said that with the utilization of Domain-Specific Modeling water utilities management and stakeholders can model the failures of water utilities using the notation of widely known risk and reliability analysis techniques. Then computer code can be generated based on these models that can be "stored" at the knowledge base of the early warning system. This type of code generation is a feasible with Domain-Specific Modeling and it can happen automatically without any intervention from the users. This can be translated in to increased productivity and to a reduction in the numbers of errors during the phases of modeling and coding.

3.2 Reasoning with Uncertain Data

Bayesian (or belief) networks [17] provide a formalism for reasoning under conditions of uncertainty. A belief network (BN) is a directed acyclic graph (DAG) consisting of a set of nodes that represent the variables of interest (e.g., the temperature of a device, a feature of an object, etc.), each having a finite set of states (or values), and a set of edges representing the probabilistic causal dependence among the variables. The nodes with edges directed into them are called child nodes, and the nodes from which the edges depart are called parent nodes. The DAG represents the structure of causal dependence between nodes and gives the qualitative part of caus-

al reasoning in a BN. The relations between variables and their corresponding states give the quantitative part, which consists of a Conditional Probability Table (CPT for short), $P(X_i \mid pa(X_i))$, attached to each node X_i with parents $pa(X_i)$ in the network's graph.
A belief network is a compact representation of the joint probability distribution $P(X_1, ..., X_n)$ over all the variables represented in the DAG.

$$P(X_1,...,X_n) = \prod_{i=1}^{n} P(X_i \mid pa(X_i))$$

EXAMPLE 1: *An example of a belief network is given in Figure 2. This belief network represents the joint probability distribution P(A,B,C,D,E) = P(A)·P(B|A)·P(C|A,B)·P(D|A)·P(E|B,D). In this case, pa(A) = Ø, pa(B) = {A}, pa(C) = {A,B}, pa(D) = {A} and pa(E) = {B,D}, respectively.*

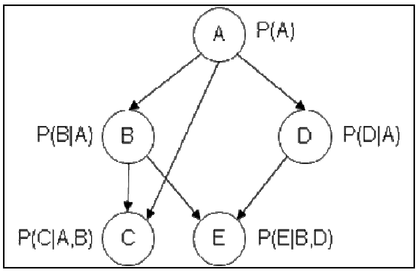

Fig. 2 A simple belief network**.**

The basic inference task of a BN, called belief updating, consists of computing the posterior probability distribution on a set of query variables *Q*, given the observation of another set of variables *E*, called the evidence (i.e., $P(Q \mid E)$). A particular instance of the above problem which has gained a lot of attention over the past decades is one in which the query set *Q* is a singleton composed of just one variable and the problem is applied to each variable of the network, except the evidence ones.
Another common inference task of a BN is finding the most probable explanation. This entails finding a variable assignment that agrees with the evidence and which has the highest probability among all such assignments.
Over the past decades, different classes of algorithms have been developed to compute the posterior probabilities or the most probable explanation. These methods fall into two general types: inference-based and search-based. Inference-based algorithms (e.g., variable elimination, tree-clustering) are better at exploiting the conditional independencies captured by the belief network, however they require exponential time and space. Search-based algorithms (e.g., depth-first search, best-first search), which

traverse a search space associated with the belief network, allow a better exploitation of the deterministic information encoded by the belief network and provide a better trade-off between time and space.

Belief networks are often used for causal representation of the phenomena involved in complex systems or processes, where information is based on expert knowledge. This approach allows for a better analysis of a dependable system as a result of additional capabilities of the belief networks with respect to the Fault Tree analysis (e.g., common cause-failure dependencies, diagnostic reasoning).

Recent work by [2; 21] has showed that it is possible and convenient to combine structural methodologies like Fault Trees with the modeling and analytical power of belief networks to provide early warnings for safety critical events, e.g., fire on ships, leakage in recovery boilers, disconnection in electrical systems, landslides, etc. Belief networks are a more general formalism than fault trees as they are able to capture several modeling aspects suitable for many applications in risk analysis and which are difficult to model within the fault tree framework. Such aspects include probabilistic gates to model common cause failures associated with some top event, or the natural ability of belief networks to use multi-valued variables to model sequentially dependent failures. In summary, belief networks provide an efficient tool to combine information from measurements signals (i.e., evidence) and calculations (i.e., belief updating) giving an early warning system that is robust to signal faults.

3.3 SaaS and BAM Technologies for Early Warning Services

A need that must be satisfied is that it must be accessed by a wide range of users in different geographical points and to allow or enhance monitoring and information flow between the actors in the wider sociotechnical system of water utilities. One way in addressing this need is by the use of state-of-the-art web technologies. A type of cloud computing architecture, known as Software as a Service (SaaS), seems to be the most appropriate alternative for providing the required early warning services to the appropriate stakeholders.

In the traditional software deployment model, businesses or end-users purchase up-front license fees and run the applications on their infrastructure. Customers host the system as well as their data on their own computers. SaaS model allows software to be deployed as a hosted service and accessed over the Internet.

In some applications domains the SaaS model has superseded the Application Service Provider model. Indeed while only one instance of an Application Service Provider application was run at a time, a single instance of a SaaS application on the other extreme can accommodate several different users sharing a single database of their data. This feature is known as multi-tenancy.

A characteristic of SaaS applications is that they have a better return on investment than their on-premises or Application Service Provider counterparts; often their services are provided for a nominal fee or sometimes free. In the case of free services, the provider usually recoups its costs from advertisement deals. Application Service Provider applications were not customizable at all, meaning that they are provided a "one-size-fits-all" version of complex applications such as enterprise resource planning, customer relationship management and so forth. On the other hand, SaaS applications are nowadays fully customizable so that users have the possibility to tailor the user interface, settings and functionalities of the SaaS application in order to meet their needs and waits.

Compared to its on-premises counterparts, SaaS presents several benefits including: reduced total cost of ownership, lower deployment, maintenance and development costs, less financial risk, easier upgrades and the lack of up-front investment. In [7], a game theoretical approach is used in order to work out the short- and long-term competition between SaaS and on-premises software providers. Furthermore, in [4] it is formally demonstrated that there is a substantial gap between the publisher's desire to invest in developing software with the SaaS deployment model rather than the on-premises version.

Monitoring of critical operation parameters in water utilities and in their wider sociotechnical system is a need that must be addressed by early warning system. In relation to this, a software technology that can be used is known as Business Activity Monitoring (BAM). The ultimate objective of this software technology is to provide real-time insight of the status and outcomes of different key business operations, processes and transactions. Such an analytical insight allows organizations to quickly spot internal issues and therefore forecast potential hitches that can reduce their outcomes. BAM importance can be explained by the increased need faced by businesses to have better business performance and an efficient measurement and analysis tool to assess their managerial tasks [13]. Businesses realize the benefits of identifying and predicting problems in order to ward off mitigating circumstances or disasters which in turn will be very costly to overcome.

BAM aims at monitoring, analyzing and alerting the output of operations within a given business. However, as businesses' size and complexity become greater and greater so do their IT requirements. Decision makers used to store all important company information and data in warehouses.

However, they eventually realized that there were several drawbacks to that approach: expensive maintenance costs, lengthy implementation cycles and the inability to handle unstructured data [10].

In our case, BAM can be used to gain insight into activities performed by users of the early warning service and thus to provide a monitoring service of important variables. Information about any unusual events occurring during a given step of water treatment will be logged and made available to decision takers and stakeholders in order to take appropriate and timely measures to avert failures and disasters.

4. Discussion and Conclusions

Waterborne disease outbreak is a major problem that is related to failures in water treatment and distribution systems. Unfortunately, the drinking water contamination problem is often detected only after a health crisis. This implies that in many cases more effective and reliable proactive strategies are needed, supported by early warning services.

In this paper we argue for a novel type of early warning service that adopts a system theoretic approach and takes into consideration the need for monitoring not only critical operational parameters in the water treatment plants but also in the components of the wider sociotechnical system.

A feasibility study has been conducted that includes an assessment of whether or not the technology need for developing such a service exists, and how existing technologies can be combined and adapted given an outline of early warning service requirements. The feasibility study takes also in to consideration the United Nations recommendations of an effective, early warning system (i.e. See section 1.1).

The results of the feasibility study are summarised in Table 3 where each selected technology is mapped with a set of stakeholders' needs and with a set of important early warning elements based on the UN recommendation for effective early warning systems.

Table 3. Mapping of selected technologies to stakeholders' needs and to elements of early warning systems.

Technology	No of Need (See Table 1)	Elements of Early Warning System (See Section 1.1)
Risk and Reliability Analysis	2, 4, 10	(i),
Domain Specific Modeling	3, 4, 5, 9	(i), (ii)
Bayesian Networks	6	(ii),
Software as a Service	1, 2, 8	(ii), (iii), (iv)
Business Activity Monitoring	3, 5, 7, 9	(ii), (iii), (iv)

For example, based on the data shown in Table 3 the Domain Specific Modelling technology can be used together with Risk and Reliability Analysis techniques in order to model effectively the knowledge about causes – failures -effects (i.e. the No 4 need listed in Table 1) and also can be used to provide the necessary element of the "knowledge of the risk".
Based on the results of the feasibility analysis it is concluded that a novel, reliable and effective web-based early warning service for water utilities can be developed by combining and adapting methods and tools derived from multiple disciplines including safety engineering, artificial intelligence, software and web engineering. This however requires the support and cooperation of all stakeholders in the wider sociotechnical system. Especially, the support of state agencies is a necessity because they are well-positioned to provide the necessary institutional support for establishing, updating and maintaining such early warning services.

Acknowledgements

This work is supported by the research project SCEWA (Grand No 2007-DRP-2-S5), funded by the Irish Environmental Protection Agency under the DERP grand scheme.

References

1. Andrews, J. D., Moss, T.R. (1003). Reliability and Risk Assessment, 1st Ed. Longman Group, UK.
2. Bobbio, A., Portinale, L., Minichinio, M., Ciancarmela, E. (2001) Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. Reliability Engineering and System Safety 71: 249-360.
3. Bradley R. (2007). Cryptosporidium-Who is responsible? Retrieved January 2009, from http://www.mlaw.ie/news/cryptosporidium-who-is-responsible
4. Choudhary, V., (2007). Software as a Service: Implications for Investment in Software Development, *Hawaii International Conference on System Sciences*, vol. 0, no. 0, pp. 209a, 40th Annual Hawaii International Conference on System Sciences (HICSS'07).
5. Corso, P. S., Kramer, M.H., Blair, K.A., Addiss, D.G., Davis, J.P., and et al. (2003). Cost of illness in the 1993 waterborne cryptosporidium outbreak, Milwaukee, Wisconsin. Emerg Infect Dis., 9(4):426:31.
6. Dokas I.M. (2008). Early Warning Systems: Views and Application Domains, (A report submitted to the Irish Environmental Protection Agency in conformity with the specifications of the research project Supporting the Concept of Early Warning Analysis – SCEWA). Cork, Ireland.

7. Fan, M., Kumar S., Whinston A. B. (2009). Short-term and long-term competition between providers of shrink-wrap software and software as a service. European Journal of Operational Research Vol 196 (2) pp. 661-671.
8. Hartl G. (2004). WHO issues revised drinking water guidelines to help prevent water-related outbreaks and disease. Retrieved January 2009, from http://www.who.int/mediacentre/news/releases/2004/pr67/en/index.html
9. Hasan, J., Goldbloom-Helzner, D., Ichida, A., Rouse, T., and Gibson. M. (2005).Technologies and techniques for early warning systems to monitor and evaluate drinking water quality: A state-of-the-art review. (Report No. EPA/600/R-05/156). U.S. Environmental Protection Agency.
10. Hellinger, M., Fingerhut, S., (2002) Business Activity Monitoring: EAI Meets Data Warehousing. eAI Journal, pp 19—21.
11. ILSI (1999). Early Warning Monitoring to Detect Hazardous Events in Water Supplies. Retrieved January 2009, from http://rsi.ilsi.org/NR/rdonlyres/C3630708-2865-429A-8965-827005579264/0/EWM.pdf
12. ISDR (2006). Global survey of early warning systems. Retrieved January 2009, from http://www.un.org/ga/61/second/ac261crp1.pdf.
13. Kang, J. G., Han, K.H. (2008). A Business Activity Monitoring System Supporting Real-Time Business Performance Management, Third International Conference on Convergence and Hybrid Information Technology, pp 473-478.
14. Kelly, S., Tolvanen, J.-P. (2008). Domain-Specific Modeling. Enabling Full Code Generation.John Wiley & Sons, Inc, Chichester.
15. Mac Kenzie, W. R., Hoxie, W,R., Proctor N,J., Gradus, M.E., Blair M.S., and et al. (1994). A massive outbreak in Milwaukee of cryptosporidium infection transmitted through the public water supply. N Engl J Med, pp. 331(3):161:7.
16. Narayanan, N.H. and Viswanadham, N. (1987). A methodology for knowledge acquisition and reasoning in failure analysis of systems. IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-17, No. 2, pp. 274-288.
17. Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann.
18. Semenza J.C., Nichols G. (2007). Cryptosporidiosis surveillance and waterborne outbreaks in Europe. Euro Surveill. 12(5):pii=711. Retrieved January 2009 from http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=711
19. Stamatelatos, M. Vesley, W. (2002). Fault tree handbook with aerospace applications, Washington: NASA.
20. U.S. EPA (2005). Watersentinel system architecture draft version 1.0. (Report No. EPA 817-D-05-003) U.S. Environmental Protection Agency.
21. Widarsson, B., Dotzauer, E. (2007). Bayesian network based early warning for leakage in recovery boilers. Applied Thermal Engineering 28: 754-760.